

CC cell viability indicates the presence of an inhibitor and a decrease in
 CC cell viability indicates the presence of an enhancer. Optionally, the
 CC method involves detecting the presence of large and small caspase
 CC subunits after contacting cell transformed with the vector expressing
 CC (1), with the candidate compound. A decrease in processing indicates the
 CC presence of an inhibitor and an increase in processing indicates the
 CC presence of an enhancer. Preferably, the large and small subunits of
 CC caspase-3, caspase-7 or caspase-9 are detected. (1) is also useful for
 CC identifying a compound that inhibits Smac binding to Smac-binding
 CC molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
 CC or a full-length IAP). (11) is useful in gene therapy techniques. The
 CC present sequence represents the amino acid sequence of Smac protein.

XX Sequence 239 AA;

Alignment Scores:

Pred. No.:	3.17e-105	Length:	239
Score:	1201.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	49.10%	Indels:	0
DB:	23	Gaps:	0

US-09-939-293-1 (1-1358) x AA078447 (1-239)

QY 20 ATGGCGCTCGAAGAGTGGTGGCGGCGAGCGTAATCTCATTTCCAGGTACAGAG 79
 DB 1 MetAlaAlaLeuLysSerTrpLeuSerArgSerValThrSerPheArgTyrArgGln 20
 QY 80 TGTGTGTGTCTCCGTGTGTGGCTAACTTTAAGAGCGGTCTTCTCAGATTGTATAGA 139
 DB 21 CysLeuCysValProValValAlaAsnPhelLysArgCysPheSerGluLeuLeuArg 40
 QY 140 CCATGGCACAACAACTGTGACGATGGCTTGGAGTACCCGTGTGGCTTCTATTGCA 199
 DB 41 ProTrpHisLysThrValThrIleGlyPheGlyValThrLeuCysAlaValProIleAla 60
 QY 200 CAGAAATCAGAGCGCTCATTCCTAGTAGTAGAGATGAGAGAGCGAGCTGCTTG 259
 DB 61 GlnLysSerGluProHisSerLeuSerSerIuAlaLeuMetArgAlaValSerLeu 80
 QY 260 GTAACAGATACACCTTACCTTCTCTCTGACACACATATGCGTGTATGAAGCTATT 319
 DB 81 ValThrAspSerThrSerThrPheLeuSerGlnThrThrTyAlaLeuIleGluAlaIle 100
 QY 320 ACTGAATATCTAAGAGCTGTTTATACCTTAACTTCTTTACCGACAATATACAGTTTA 379
 DB 101 ThrGluTyrThrLysAlaValLysThrLeuThrSerLeuTyrArgGlnTyrThrSerLeu 120
 QY 380 CTTCGGAAATGAAATTCAGAGAGAGATGAAGTGTGGCAGGTGATATAGAGCCAGA 439
 DB 121 LeuLysLysMetAsnSerGluGluGluAspIuValTrpGlnValIleIleGlyAlaArg 140
 QY 440 GCTAGATGACTTCAAAAACACCAAGAGTACTTGAAGCTGGAACCACTTGGATGCTGCA 499
 DB 141 AlaGluMetThrSerLysHisGlnGluTyrLeuLysLeuGluTyrThrThrPheThrAla 160
 QY 500 GTTGCTCTTTCAGAGATGGCAGCAGAACTCATATCAAACTGGCGCAGATCAGGCCCTCT 559
 DB 161 ValGlyLeuSerGluMetAlaAlaGluValAlaLysGlnThrGlyAlaAspGlnAlaSer 180
 QY 560 ATAACCCGCGAGATCATCTCAGCTGGTGAACCTGCAGGTGGGAAGAGGTCCACCAAGTTC 619
 DB 181 IleThrAlaArgAsnHisIleGlnLeuValLysLeuGlnValIleGluValHisGlnLeu 200
 QY 620 TCCCGAAGAGCAGAAACCACTGGCAGACACATAGATAGAAGCTCCGTCAGAAACA 679
 DB 201 SerArgLysAlaGluThrLysLeuAlaGluAlaGlnIleGluGlnLeuArgGlnLysThr 220
 QY 680 CAGAGGAAGGGGAGGAGCGGGCTGAGTGGAGCAGAGAGAGCGCTACCTGCTGAGAGAT 736
 DB 221 GlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 239

RESULT 3

AAB54139
 ID AAB54139 standard; Protein: 227 AA.

XX AAB54139;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:591.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neotrophic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.

KW Homo sapiens.

PN WO200055320-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05989.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI: 2000-579444/54.

DR N-PSDB; AAC98904.

PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 PS Claim 11; Page 1027-1028; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neotrophic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used for design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 227 AA;

Alignment Scores:

Pred. No.:	2.98e-99	Length:	227
Score:	1138.00 <td>Matches:</td> <td>226</td>	Matches:	226
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	46.52%	Indels:	0
DB:	21	Gaps:	0

CC The invention relates to an isolated Smac peptide or polypeptide (I)
CC and an isolated nucleic acid (II) encoding (I). Also described is a
CC method of identifying a compound that inhibits apoptosis, comprising:
CC (a) separately contacting several cell populations expressing a
CC cytosolic Smac (a Smac isoform that begins with MKSPDYF sequence,
CC replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
CC and residues 56-60 of (I)) and an inhibitor of Bid (Bcl2 interacting
CC domain) with a compound to be tested for apoptotic inhibiting activity;
CC (b) incubating the cell populations with a direct stimulus of the cell
CC death pathway; and (c) measuring the specific apoptotic activity of the
CC cell populations, where inhibition of the specific apoptotic activity is
CC indicative that the compound is an inhibitor of apoptosis. (I) and (II)
CC are useful for inducing apoptosis in a cell. The Smac polypeptide and
CC polynucleotide are useful for stimulating apoptosis in a neoplastic or
CC tumour cell which overexpresses an inhibitor of caspase-3, where the
CC inhibitor inhibits activation or activity of caspase-3, caspase-7 or
CC caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
CC (I) is useful for identifying an inhibitor or enhancer of a caspase-
CC mediated apoptosis which involves contacting a cell transformed or
CC transfected with a vector expressing (I) with a candidate inhibitor or
CC candidate enhancer; and detecting cell viability, where an increase in
CC cell viability indicates the presence of an inhibitor and a decrease in
CC cell viability indicates the presence of an enhancer. Optionally, the
CC method involves detecting the presence of large and small caspase
CC subunits after contacting cell transformed with the vector expressing
CC (I), with the candidate compound. A decrease in processing indicates the
CC presence of an inhibitor and an increase in processing indicates the
CC presence of an enhancer. Preferably, the large and small subunits of
CC caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
CC identifying a compound that inhibits Smac binding to Smac-binding
CC molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
CC or a full-length IAP). (II) is useful in gene therapy techniques. The
CC present sequence represents the amino acid sequence of Smac mutant
CC Smac-N39.
XX
SQ Sequence 39 AA:

Alignment Scores:
Pred. No.: 3,71e-09 Length: 39
Score: 186.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 7.60% Indels: 0
DB: Gaps: 23

US-09-939-293-1 (1-1358) x AAU78436 (1-39)

QY 185 GCGGTTCCTATTGCACAGAAATCAGAGCCTCATTCCTTACTAGTGAAGCATGTGATGAGG 244

Db 1 AlavAlProIleAlaGlnLySerGluProHisSerLeuSerSerGluAlaLeuMetArg 20

QY 245 AGACAGAGTCTTTGGTGAACAGATAGCACCCTTCTCTCTCAGACACATAT 301

Db 21 ArgAlaValSerLeuValThrAspSerThrSerThrPheLeuSerGlnThrThrTyr 39

RESULT 7

AAU78439
ID AAU78439 standard; Peptide: 35 AA.

AC AAU78439;

DT 18-JUN-2002 (first entry)

DE Inhibitor of apoptosis (IAP) protein Smac, peptide Smac-N35.

KW Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1, BIR2;
KM Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;

OS neoplastic cell; tumour.

XX Homo sapiens.

XX MO200216418-A2.
XX PN

PD 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26492.
XX
XX 24-AUG-2000; 2000US-227735P.
XX
XX (UYJE-) UNIV JEFFERSON THOMAS.
XX
XX Alnemr1 ES;
XX
XX WPI; 2002-304115/34.
DR

PT Novel Smac peptides and polynucleotides encoding the peptides, useful
PT for stimulating apoptosis in neoplastic or tumour cell which
PT overexpresses inhibitor of caspase, and for identifying apoptosis
PT modulating compounds

PS Example 4; Page 47; 78pp; English.

XX The invention relates to an isolated Smac peptide or polypeptide (I)
XX and an isolated nucleic acid (II) encoding (I). Also described is a
XX method of identifying a compound that inhibits apoptosis, comprising:
XX (a) separately contacting several cell populations expressing a
XX cytosolic Smac (a Smac isoform that begins with MKSPDYF sequence,
XX replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
XX and residues 56-60 of (I)) and an inhibitor of Bid (Bcl2 interacting
XX domain) with a compound to be tested for apoptotic inhibiting activity;
XX (b) incubating the cell populations with a direct stimulus of the cell
XX death pathway; and (c) measuring the specific apoptotic activity of the
XX cell populations, where inhibition of the specific apoptotic activity is
XX indicative that the compound is an inhibitor of apoptosis. (I) and (II)
XX are useful for inducing apoptosis in a cell. The Smac polypeptide and
XX polynucleotide are useful for stimulating apoptosis in a neoplastic or
XX tumour cell which overexpresses an inhibitor of caspase, where the
XX inhibitor inhibits activation or activity of caspase-3, caspase-7 or
XX caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
XX (I) is useful for identifying an inhibitor or enhancer of a caspase-
XX mediated apoptosis which involves contacting a cell transformed or
XX transfected with a vector expressing (I) with a candidate inhibitor or
XX candidate enhancer; and detecting cell viability, where an increase in
XX cell viability indicates the presence of an inhibitor and a decrease in
XX cell viability indicates the presence of an enhancer. Optionally, the
XX method involves detecting the presence of large and small caspase
XX subunits after contacting cell transformed with the vector expressing
XX (I), with the candidate compound. A decrease in processing indicates the
XX presence of an inhibitor and an increase in processing indicates the
XX presence of an enhancer. Preferably, the large and small subunits of
XX caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
XX identifying a compound that inhibits Smac binding to Smac-binding
XX molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
XX or a full-length IAP). (II) is useful in gene therapy techniques. The
XX present sequence represents the amino acid sequence of Smac peptide
XX Smac-N35.
XX

SQ Sequence 35 AA:

Alignment Scores:
Pred. No.: 4.36e-07 Length: 35
Score: 164.00 Matches: 35
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 6.70% Indels: 0
DB: Gaps: 23

US-09-939-293-1 (1-1358) x AAU78439 (1-35)

QY 185 GCGGTTCCTATTGCACAGAAATCAGAGCCTCATTCCTTACTAGTGAAGCATGTGATGAGG 244

Db 1 AlavAlProIleAlaGlnLySerGluProHisSerLeuSerSerGluAlaLeuMetArg 20

QY 245 AGACAGAGTCTTTGGTGAACAGATAGCACCCTTCTCTCTC 289

Db 21 ArgAlaValSerLeuValThrAspSerThrSerThrPheLeuSer 35

RESULT 8
 AAU78435
 ID AAU78435 standard; Peptide: 30 AA.
 XX
 AC AAU78435;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Inhibitor of apoptosis (IAP) protein Smac, mutant Smac-N30.
 XX
 KW Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1; BIR2;
 KW Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;
 KW neoplastic cell; mutant; tumour.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200216418-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26492.
 XX
 PR 24-AUG-2000; 2000US-227735P.
 XX
 PA (UYJE-) UNIV JEFFERSON THOMAS.
 XX
 PI Alnemrl ES;
 XX
 DR WPI: 2002-304115/34.
 XX
 PT Novel Smac peptides and polynucleotides encoding the peptides, useful
 PT for stimulating apoptosis in neoplastic or tumour cell which
 PT overexpresses inhibitor of caspase, and for identifying apoptosis
 PT modulating compounds -
 XX
 PS Example 3: Fig 7; 78pp; English.
 XX
 CC The invention relates to an isolated Smac peptide or polypeptide (I)
 CC and an isolated nucleic acid (II) encoding (I). Also described is a
 CC method of identifying a compound that inhibits apoptosis, comprising:
 CC (a) separately contacting several cell populations expressing a
 CC cytosolic Smac (a Smac isoform that begins with MKSDPYF sequence,
 CC and residues 56-60 of (I)) and an inhibitor of BID (bcl2 interacting
 CC domain) with a compound to be tested for apoptotic inhibiting activity;
 CC (b) incubating the cell populations with a direct stimulus of the cell
 CC death pathway; and (c) measuring the specific apoptotic activity of the
 CC cell populations, where inhibition of the specific apoptotic activity is
 CC indicative that the compound is an inhibitor of apoptosis. (I) and (II)
 CC are useful for inducing apoptosis in a cell. The Smac polypeptide and
 CC polynucleotide are useful for stimulating apoptosis in a neoplastic or
 CC tumour cell which overexpresses an inhibitor of caspase, where the
 CC inhibitor inhibits activation or activity of caspase-3, caspase-7 or
 CC caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
 CC (I) is useful for identifying an inhibitor or enhancer of a caspase-
 CC mediated apoptosis which involves contacting a cell transformed or
 CC transfected with a vector expressing (I) with a candidate inhibitor or
 CC candidate enhancer; and detecting cell viability, where an increase in
 CC cell viability indicates the presence of an inhibitor and a decrease in
 CC cell viability indicates the presence of an enhancer. Optionally, the
 CC method involves detecting the presence of large and small subunits
 CC subunits after contacting cell transformed with the vector expressing
 CC (I), with the candidate compound. A decrease in processing indicates the
 CC presence of an inhibitor and an increase in processing indicates the
 CC presence of an enhancer. Preferably, the large and small subunits of
 CC caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
 CC identifying a compound that inhibits Smac binding to Smac-binding
 CC molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
 CC or a full-length IAP). (II) is useful in gene therapy techniques. The
 CC present sequence represents the amino acid sequence of Smac mutant
 CC Smac-N30.

XX
 SO Sequence 30 AA;
 Alignment Scores:
 Pred. NO.: 6.27e-05 Length: 30
 Score: 141.00 Matches: 30
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.76% Indels: 0
 DB: 23 Gaps: 0
 US-09-939-293-1 (1-1358) x AAU78435 (1-30)
 QY 185 GCGGTCCTATTGGCAGCAATACAGACCTCATTCCTTAGTGAACCATTTGAGC 244
 Db 1 AaValProlleAlaGlnLysSerGIupRohIsSerSerSerGIuAlaIeuMetArg 20
 QY 245 AGAGCAGTGTCTTGTGTAACAGATGACCC 274
 Db 21 ArGaIaValSerIeuValThrAspSerThr 30
 RESULT 9
 AAM52315
 ID AAM52315 standard; Protein: 502 AA.
 XX
 AC AAM52315;
 XX
 DT 18-JAN-2002 (first entry)
 XX
 DE Human WASP protein.
 XX
 KW Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;
 KW metastatic cancer; parasitic infection; cytotoxic; Human; WASP.
 XX
 OS Homo sapiens.
 XX
 PN WO200171356-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-FR00843.
 XX
 PR 22-MAR-2000; 2000FR-0003637.
 XX
 PA (CNRS) CENT NAT RECH SCI.
 PA (CURRI-) INST CURIE.
 PI Friedlizi J, Friederich E, Golsteyn RM, Louvard D, Nolreaux V;
 SY Sykes C;
 DR WPI: 2001-639148/73.
 XX
 PT Identifying modulators of actin polymerization, potentially useful for
 PT treating tumor metastasis and parasitic infection, using proteins that
 PT contain Ena/VASP binding sites -
 XX
 PS Claim 13; Pages 95-96; 109pp; French.
 XX
 CC The present invention relates to a method for identifying modulators of
 CC actin polymerisation. The method involves using proteins that contain at
 CC least one binding motif for proteins of the Ena/VASP
 CC (vasodilator-stimulated phosphoprotein) family in the preparation of
 CC reagents for identification/screening of molecules that modulate
 CC formation of the actin cytoskeleton. The proteins used in the method
 CC (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not
 CC bind to the Arp2/3 protein complex. The modulators identified by the
 CC method are potentially useful for treating disorders of actin
 CC polymerisation, e.g. metastatic cancer or parasitic infection; and as
 CC cytotoxic agents. The present sequence one such protein with binding
 CC motif(s) for Ena/VASP proteins, which was used in the method of the
 CC present invention.
 XX
 SO Sequence 502 AA;


```

QY 1075 TCCTGAACTTAAGGCAATGACAAAAGACCTCTCTCTGTACCCAGTAGGCAAAATGCC 1016
      :::::
Db 270 LeuphieserAraAlaGlyIleSerIu-----AlaGlnLeu 281
QY 1015 TTTGGGTGAGCTAAATAAATGGTAAAGACACCTGTACAGTGGGCT----- 965
      |||
Db 282 ThrAspAlaGluThrSerLysLeuIleTyrAspPheIleGluAspGlnGlyIleuGlu 301
      |||
QY 964 -----GAAATGTTAAACAGGCTGCAGTCCCA-----AGG 935
      |||||
Db 302 AlaValArgGlnGluMetArgArgGlnGlnProLeuProProProProProSerArg 321
QY 934 GCTAAGACACAGGTCCAGGCGCAAGCTGAGACACAGAGACGACATCAGACGTCACAAAG 875
      |||||
Db 322 GlyGlyAsnGlnLeuProArg---ProProlIleValGlyGlyAsnLys----- 336
QY 874 GGTCTGCGCTGATGTGGCCAGGCGAGGACCT-----GCCGCC 839
      |||||
Db 337 -----GlyArgSerGlyProLeuProProValProLeuGlyIleAlaPro 351
QY 838 TCTTCTCGGTGACAGACAGTATGCCAACCTGGGACGGTGGCATCTGCCCTGCTTT 779
      |||||
Db 352 ProProProThrProArgGlyProProProProGlyArgGlyGlyProProProPro 371
QY 778 CCCCACTGAGTGGGAGACAGGAGAGTGTGCTCAGGCGCCCTCAATGCTCAGCAGGTAGGC 719
      |||||
Db 372 ProProAlaThrGlyArgSerGlyProLeuProProProPro--ProGlyAlaGlyGlyP 391
QY 718 CTCCTGCTCCGACATGACCCGCTCCCTCCCTGCTGCTGTTTCTGACGAGGCTCTTC 659
      |||||
Db 391 roPMePProProProProProProProProProProSerSerGly----- 407
QY 658 TATCTGTGCTTGTGCGACGCTTGTCTTCCGCGGAGACCTGTGACACTCTTCCAC 599
      |||||
Db 408 -----AsnGlyProAlaProProProLeuProP 417
QY 598 CTGCACTTTCACACCTGTAAT---GTGATTCCTGGCGGTTATAGAGACCTGATC 548
      |||||
Db 417 roAlaLeuValProAla--GlyGlyLeuAlaProGlyGlyArgGlyAlaLeu 434
      |||||

RESULT 11
AAU79074
ID AAU79074 standard; Protein; 502 AA.
AC
XX AAU79074;
XX
XX 02-JUL-2002 (first entry)
XX
XX Human Wiscott-Aldrich Syndrome protein, WASP.
XX
XX Biosensor: synthon; aminoxy group; GTP-activated Rho GTPase;
XX Wiscott-Aldrich syndrome; calmodulin; rac; cdc42; MAP kinase; Erk1; Erk2;
XX Erk3; Erk; IGE receptor; F-cepSimonRI actin; alpha-actin; myosin;
XX major histocompatibility protein; WASP.
XX
XX Homo sapiens.
XX
XX WO200208245-A2.
XX
XX 31-JAN-2002.
XX
XX 13-JUL-2001; 2001WO-US22194.
XX
XX 13-JUL-2000; 2000US-218113P.
XX 29-SEP-2000; 2000WO-US26821.
XX 28-MAR-2001; 2001US-279302P.
XX 20-APR-2001; 2001US-0839577.
XX
XX (SCRI ) SCRIPPS RES INSTR.
XX
XX Hahn KM, Toutchkine A, Muthyala R, Kraynov V, Bark SJ, Burton DR,
XX Chamberlain C;

```

```

XX DR WPI: 2002-315250/35.
XX
XX PT Novel synthetic intermediate compounds useful for preparing modified
XX peptides and peptide conjugates which are useful in biosensors for
XX detecting the location of cellular proteins in living cells -
XX
XX PS Example 8; Page 86; 158pp; English.
XX
XX CC The invention relates to a compound which is a synthetic intermediate
XX CC (a synthon) useful for preparing modified peptides. The compound
XX CC has the structure and formula given in the specification. Also
XX CC included are a peptide (Iia) comprising a backbone and one or more
XX CC aminoxy groups, provided the peptide is not glutathione and the peptide
XX CC has at least one aminoxy group that is not part of a group NH2-O-CH2-C
XX CC (=O)- positioned at the N-terminus of the peptide or that is not part of
XX CC a group -C(N-O-CH2-C(=O))- that is in the backbone, a peptide (Iib)
XX CC comprising a backbone and one or more secondary aminoxy groups, provided
XX CC the peptide having an aminoxy group that is not part of an oxime
XX CC (C=N-O-) in the backbone, identifying (IV) an optimal position for
XX CC replacement of a functional molecule on a peptide having a peptide
XX CC backbone and a known activity, by making a series of peptide conjugates,
XX CC each peptide conjugate having the same amino acid sequence and the same
XX CC functional molecule, where the functional molecule is linked at a
XX CC different location along the backbone of every peptide conjugate in the
XX CC series, and observing which functional molecule location does not
XX CC substantially interfere with the known activity of the peptide, a
XX CC polypeptide biosensor (VI) comprising a polypeptide capable of
XX CC binding a GTP-activated Rho GTPase protein, where the polypeptide is
XX CC operatively linked to a functional molecule, a fluorescent compound
XX CC (VIII) of a formula given in the specification, a peptide biosensor
XX CC comprising (VIII) linked to a protein, polypeptide, peptide, antibody or
XX CC its fragment, or a nucleic acid, attaching a biosensor to a cellular
XX CC protein within a living cell, by providing the living cell with a
XX CC biosensor capable of binding to a tag on the cellular protein, where the
XX CC tag is a peptide segment that has been fused to the cellular protein
XX CC expressed by the living cell. The synthon is useful for preparing
XX CC modified peptides. Peptides (Iia) and (Iib) are useful for preparing a
XX CC peptide conjugate comprising a peptide linked to a functional molecule.
XX CC The biosensor is useful for detecting GTP activation of a Rho GTPase
XX CC protein in a living cell. Cellular proteins which may be detected
XX CC are calmodulin, Rho GTPase, rac, cdc42, MAP kinase, Erk1, Erk2,
XX CC Erk3, Erk4, IGE receptor (F-cepSimonRI) actin, alpha-actin, myosin or a
XX CC major histocompatibility protein. Synthons and the peptides
XX CC are useful for labeling antibodies, antigens and other polypeptides.
XX CC The peptide is useful as synthetic intermediate that can be modified to
XX CC provide related peptides having altered biological, chemical or
XX CC physical properties for e.g. a peptide linked to a fluorescent label.
XX CC The biosensor is useful for diagnostic (e.g. of Wiscott-Aldrich
XX CC syndrome) or imaging purposes in vivo, within a mammalian subject. The
XX CC biosensor retains high fluorescence emission after conjugation to other
XX CC molecules and avoids problems of aggregation and insolubility. The
XX CC present sequence is the Human Wiscott-Aldrich Syndrome protein
XX CC (WASP) whose Cdc42-binding fragment is used to make a biosensor for
XX CC Cdc42.
XX
XX SQ Sequence 502 AA:
XX
XX Alignment Scores:
XX Pred. No.: 0.00562 Length: 502
XX Score: 125.50 Matches: 74
XX Percent Similarity: 35.00% Conservative: 24
XX Best Local Similarity: 26.43% Mismatches: 102
XX Query Match: 5.20% Indels: 80
XX DB: 23 Gaps: 13
XX
XX US-09-939-293-1 (1-1358) x AAU79074 (1-502)
QY 1291 CCTTCTCTAGTAGTAAATAGGTTT---TTCACCTCTTGCCCTCAGCTGTCTCAGAGAC 1235
      |||
Db 205 ProAspIleThrSerSerArgGlyLeuProAlaProGlyProSerProAlaAsp 224
      |||
QY 1234 -----AGTGGGGGCGAGATCAGAGACACATCAGAAATACATCAAGAAATCGTAC 1184

```



```

Db 98 -----ThrAlaArgGlyProArgArgValSerArgArgLeuProProGlnHis 113
OY 1032 CCAGGT-----AGGCAAAATGCTTTGGGTGTGAGGTAAATAATGGGTA 988
XX ||||| ||||| ||||| |||||
Db 114 ProGlyProArgGlyArgArgArgProGlyAlaGlyAlaGlyAlaProArgArgGly 133
OY 987 AGAGACCTGTACAGAGTGGGTGAATGTAAACAGGTGTCAGTCCCAAGGGCTAAGA 928
XX ||||| ||||| ||||| |||||
Db 134 ArgAlaArgGlyGlnAlaGlyLeu-----LeuGlyArgGlnGlyGlnGlyArg 150
OY 927 ACCAGGTCCAGGCCAAGCTGACACCAAGAGGAGCTACAGCTCAAAAGGGCTCG 868
XX ||||| ||||| ||||| |||||
Db 151 -----GlyAlaGluArgGluArgAlaLeuGlnAlaArgArgGlyArgArg 166
OY 867 CCGATTGGCCAGGAGGAGACCTGCGCCCTCTCTGCGACAGACAGCATGACCAAC 808
XX ||||| ||||| ||||| |||||
Db 167 Pro-----GlyProGluProAspGlnSerCys 175
OY 807 CTGGGAGGGTGGCATCTGCCCTGCTTTCCACACTAGTGGGAGACAGGAGTGTGC 748
XX ||||| ||||| ||||| |||||
Db 176 GlyGlyArgProArgArgAlaAlaAlaPro-----GlyArgAlaPro--- 190
OY 747 TCAGGCCCTCAATCTCAGCCAGGTAGGCT----- 717
XX ||||| ||||| ||||| |||||
Db 191 AlaAspProGln-ProProAlaProArgProAlaProAlaProAspValaArgProAla 210
OY 716 -----CGCTGCCGACT---CAGCCCGCTCCCTCCCTCCCTGTTTCTGACGG 667
XX ||||| ||||| ||||| |||||
Db 210 aaAspAlaProAlaProAlaProAlaProAlaProProProProHisLeu-----G 228
OY 666 AGCTCTCTATCT-----GTGCTTCTGCGACGCTGTTCTTCTGCT 628
XX ||||| ||||| ||||| |||||
Db 228 yAlaLeuThrAlaGlySerGlyGluArgGlnSerGlnProArgAlaGluThrLeuAr 248
OY 627 TTCGGGAGAGCTGTGTCACCTTCCACCTGCAGTTTCA-----CCAGCTGA 580
XX ||||| ||||| ||||| |||||
Db 248 gLeuGlyArg---GlyAlaProLeuProProArgAlaGluArgGlyGlyArgProLysG 267
OY 579 ATGTGATTCTCTGGGGTTATAGAGGCT 552
Db 267 nAlaGluGlnGlnGlnProLysArgPro 276

RESULT 13
AAM40792
ID AAM40792 standard; protein; 317 AA.
XX
AC AAM40792;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5723.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
PF 21-JAN-2000; 2000US-0488725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PR 09-JUL-2000; 2000US-0598042.
XX
PR 19-JUL-2000; 2000US-0620312.
XX
PR 03-AUG-2000; 2000US-0653450.
XX
PR 14-SEP-2000; 2000US-0662191.

```

```

PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AAI59948.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5723; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemia and
XX C.N.S disorders.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 317 AA;
XX
XX
Alignment Scores:
Pred. No.: 0.0174 Length: 317
Score: 119.50 Matches: 82
Percent Similarity: 36.90% Conservative: 25
Best Local Similarity: 28.28% Mismatches: 99
Query Match: 4.95% Indels: 84
DB: 22 Caps: 17

US-09-939-293-1 (1-1358) x AAM40792 (1-317)
OY 1320 TTGAGAGCACCAGCTAAACACTGTGCACCCCTCTCTAGTAGTAATAGTTTTCATCTC 1261
XX ||||| ||||| ||||| |||||
Db 38 ValGlnAlaProGly-----AlaPro-----GlySer 46
OY 1260 CTTCGGCTCAGC-----TTCCTTCACAGGACGTGGGGCGCATCAAGAAC 1213
XX ||||| ||||| ||||| |||||
Db 47 LeuGlyValSerHisAlaAlaAlaProProAlaArgProGlnGlyAlaAlaGlnSerPro 66
OY 1212 CATGCAAAATACATACAAAGAAATCGTAAACATGAGACAGTCCCTCCCTGCCACCA 1153
XX ||||| ||||| ||||| |||||
Db 67 HisArgGly-----ArgArgHisGlyGlyGlyAlaGlyLeuProProProArgSer 84
OY 1152 ACTGGCATCCCAACAGAGGGAACAAGTACTAAATCATTTTTCAGCAGCTAAATAGACTG 1093
XX ||||| ||||| ||||| |||||
Db 85 ProArgPheProGlnGlnSerValProAlaSerThrSer----- 97
OY 1092 AAAACAGGTTAAACAGATGCTGAACCTTAAGGCGCATGACAAAAGACTCTCTCTGAC 1033
XX ||||| ||||| ||||| |||||
Db 98 -----ThrAlaArgGlyProArgArgValSerArgArgLeuProProGlnHis 113
OY 1032 CCAGGT-----AGGCAAAATGCTTTGGGTGTGAGGTAAATAATGGGTA 988
XX ||||| ||||| ||||| |||||
Db 114 ProGlyProArgGlyArgArgArgProGlyAlaGlyAlaGlyAlaProArgArgGly 133
OY 987 AGAGACCTGTACAGAGTGGGTGAATGTAAACAGGTGTCAGTCCCAAGGGCTAAGA 928
XX ||||| ||||| ||||| |||||
Db 134 ArgAlaArgGlyGlnAlaGlyLeu-----LeuGlyArgGlnGlyGlnGlyArg 150

```

```
OY 927 ACCAGTCCAGCGCAGCCTAGACACAGGACCTCAGCTCACAAGGCGCTCG 868
Db 151 -----GlyAlaGluArgGluAlaLeuGlnAlaArgGlyArg 166
OY 867 CCTGATTTGGCAGGCGAGCAGCTGGCGCTTCTCCGTGCACAGACAGTATGCCAAC 808
Db 167 Pro-----GlyProGluProAspGlnSerCys 175
OY 807 CTGGGCGAGGTGGATCTGCCCCCTTCCCATGAGTGGAGAGAGGCGAGTGTGC 748
Db 176 GlyIleArgProArgArgAlaAlaAlaPro-----GlyArgAlaPro 190
OY 747 TCAGGCGCTCAATCTCAGCAGAGTAGGCT----- 717
Db 191 AlaAspProGln-ProProAlaProArgProAlaProAlaProAlaProAla 210
OY 716 -----CTGCTCCGACT--CAGCCGCTCCTCCCTTCCCTGTTTTCGACGG 667
Db 210 aAspAlaProAlaProAlaProAlaProAlaProProProProHisIleu-----GI 228
OY 666 AGCTCTCTATCT-----GTGCTCTGCCAGCTTGTTCTGCT 628
Db 228 yAlaIleuThrAlaGlySerGlyGluArgGlnSerGlnProAlaGluThrIleuAr 248
OY 627 TTCCGGAGAGCTGTGCACCTCTCCACCTGCAGTTCA-----CCAGCTGA 580
Db 248 gIeuGlyArg--GlyAlaProIleuProProAlaGluArgGlyGlyArgProLysGI 267
OY 579 ATGTGATCTCTGGCGTTATAGAGCCT 552
Db 267 nAlaGluGlnGlnInProLysArgPro 276

RESULT 14
ABG03762
IT ABG03762 standard; Protein; 276 AA.
XX
AC ABG03762;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #3753.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB: AAS67949.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
PS Claim 20; SEQ ID No 34121; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
```

```
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
```

SQ Sequence 276 AA;

Alignment Scores:

Pred. No.:	0.0354	Length:	276
Score:	116.00	Matches:	65
Percent Similarity:	35.10%	Conservative:	21
Best Local Similarity:	26.53%	Mismatches:	74
Query Match:	4.81%	Indels:	85
DB:	22	Gaps:	9

US-09-939-293-1 (1-1358) x ABG03762 (1-276)

```
OY 1161 CCTGCCACAACCTGACATCCACAGAG-----GGAACAGACT 1123
Db 79 ProAlaProSerGlyIleuProSerGluArgLysArgAspAlaAlaAlaIleuSer 98
OY 1122 AAATCATTTTGAACGACGTAATTAAGACTGATAAAGGTTAAACAGTTGCTGAACTTAAG 1063
Db 99 AlaSerAlaIleuThr----- 103
OY 1062 GGCATGCAAAAAGAGACTCTCTCTGACCCAGTAGCAAAATGCTTGGTGAGG 1003
Db 104 GlyIleuThrLysArgProIleuSerSerThrProIleuSerAlaLeuGly----- 121
OY 1002 TAAAAAATGGGTAGAGCAGCTGTACAGAGTGGGTTGAATGTTAAACAGGTCAGT 943
Db 121 ----- 121
OY 942 GCCCAAGGCTTAAGAACCAAGGTCAGCGCAAGCTTAGACACAGAGGCACTCAGAGT 883
Db 122 -----ArgIleuAlaGluAlaAlaValAla 129
OY 882 CACAAA---GGCGTCGGCGTGAATTGGCCAGGCGAGGACTGCCGCTTCTCGGTGA 826
Db 130 GluLysArgAlaIleuSerProSerIleLysGluProSerValValProIle----- 146
OY 825 CAGACAGTATGCCAAC-----CTGGCAGGAGTGGCACTGTGCGCTTCCCACT 772
Db 147 ---GluValIleuProThrValIleuAspGluIleGluAlaAlaIleuSerThrProAlaThr 165
OY 771 GAGTGGGAGACAGGCGAGTGTCTCAGGCGCTCAATCTCAGCAGGAGTAGGCTCTCGC 712
Db 166 MetThrGlySerArgAlaCysAlaGlyPro***SerSer***ArgSerProAlaPro 185
OY 711 TCCGAGTACGCCCGCTCCCTCTCTCTCTGTTTGTGACGAGCTCT-----TCT 658
Db 186 SerIleuThrAlaProSerThr***AlaSerCysThrTrpProArgSerSerProThrSer 205
OY 657 ATCTGTGCTTCCAGCAGTGTGTTCTTCTCCGGAAGAGTGGGCACTCTTCCAC 598
Db 206 SerProIleuArgAlaSerIleuArgIleuCysValAlaSerCysGlyGly---ThrProTr 224
OY 597 TGCAGTTTCACCAAGCTGAATGTGATTCCTGCGGCTTATAGAGGCTGATCTCGCCAGTT 538
```


Search completed: February 20, 2003, 14:57:10
Job time : 81.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 20, 2003, 14:53:58 ; Search time 45 Seconds

(without alignments)
5802.246 Million cell updates/sec

Title: US-09-939-293-1

Perfect score: 2446

Sequence: 1 ggcgtccgcgcgcgcacacaa.....ggacttaacacagaagaaaaa 1358

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_n2p.model -DEV=xlp
-O=/cgn2.1/USPTO-spool/US09939293/runat_20022003.111511.14892/app-query.fasta.1.1543
-DB=PIR.73 -QFMT=fastan -SUFIX=xrpr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blissum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=Pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09939293_ecgn_1.1.73_etunal.20022003.111511.14892 -NCPD=6 -ICPU=3
-NO_XLPTX -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR.73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	125.5	5.2	805	2 A55197	Wiskott-Aldrich sy
C 2	123	5.1	802	2 T49385	hypothetical prote
C 3	116.5	4.8	1133	2 T22976	hypothetical prote
C 4	112	4.6	3660	1 S02041	dystrophin, muscle
C 5	111	4.6	1323	2 I78557	N-methyl-D-asparta
C 6	111	4.6	1356	1 C45219	N-methyl-D-asparta
C 7	107.5	4.4	1396	2 T10627	hypothetical prote
C 8	107.5	4.4	1520	2 T23620	hypothetical prote
C 9	106.5	4.4	3164	1 MMBE6	hypothetical prote
C 10	105.5	4.3	4574	2 G02520	ULB6 protein - hum
C 11	105	4.4	337	2 C90968	plectin - human
C 12	105	4.3	791	2 H72552	Espr-like protein
C 13	104.5	4.3	517	2 T10927	hypothetical prote
C 14	104.5	4.3	581	2 E75383	3C3.18c protein - conserved hypothet

15	104.5	4.3	1220	2 A56136	jagged protein pre
16	103.5	4.2	345	2 T08700	hypothetical prote
17	103.5	4.2	4684	2 A59404	plectin (imported)
C 18	103	4.3	420	2 T46910	hypothetical prote
19	103	4.2	592	2 B48315	lamin B2 - mouse
20	102.5	4.2	729	2 E81916	probable p130cas
21	102.5	4.2	3924	2 S37431	ankyrin 2, neurom
22	102.5	4.2	5376	2 T42215	zonadhesin - mouse
23	102	4.2	508	1 KRSH12	keratin type II, m
24	102	4.2	2116	2 A26655	myosin heavy chain
25	101.5	4.1	373	2 T34743	hypothetical prote
26	101.5	4.1	1214	2 T00356	hypothetical prote
27	101	4.1	859	2 T26134	hypothetical prote
28	100.5	4.1	741	2 S39082	myosin heavy chain
29	100.5	4.1	1084	2 T08583	cellulose synthase
30	100	4.1	640	2 T08179	LRG3 protein - Chl
31	100	4.1	701	2 H98120	choline binding pr
32	100	4.1	4377	2 A55575	ankyrin 3, long sp
33	99.5	4.1	955	2 S24348	myosin heavy chain
34	99.5	4.1	1087	1 OFMSH	neurofilament trip
C 35	99.5	4.1	1213	2 A41724	limb deformity (ld
C 36	99	4.1	616	2 I38155	DNA-binding regula
C 37	99	4.1	660	2 JW0067	chitinase (EC 3.2.
C 38	99	4.1	721	2 E70766	hypothetical prote
C 39	99	4.1	1104	2 S59310	probable membrane
C 40	98.5	4.1	988	2 S37078	chloride channel p
41	98	4.0	548	1 OFPGL	neurofilament trip
42	98	4.0	1072	1 A37221	polyketide synthas
C 43	98	4.0	4427	2 PN0637	hypothetical prote
C 44	97.5	4.0	374	2 T16486	protein T04A6.13.1
45	97.5	4.0	762	2 G88436	

ALIGNMENTS

RESULT 1

A55197 Wiskott-Aldrich syndrome protein WASP - human

C:Species: Homo sapiens (man)

C:Date: 23-Mar-1995 #sequence_revision 24-Nov-1999 #text_change 24-Nov-1999

C:Accession: A54747; A55197; I38931

R:Derry, J.M.J.; Ochs, H.D.; Francke, U.

Cell 78, 635-644, 1994

A:Title: Isolation of a novel gene mutated in Wiskott-Aldrich syndrome.

A:Reference number: A54747; MUID:94349367; PMID:8069912

A:Accession: A54747

A:Molecule type: mRNA

A:Residues: 1-328, 'A', 330-366, 'LHNRPLQDLVDLHCRPLRLVGRCHN', 'HNRHRRRPARGMDFPLPSLT

A:Cross-references: GB:U12707

A:Note: this sequence is corrected in reference A55197

Cell 79, 922a, 1994

A:Reference number: A55197

A:Contents: erratum

A:Accession: A55197

A:Molecule type: mRNA

A:Residues: 1-424, 'PG', 427-502 <DER>

A:Cross-references: GB:U12707; NID:9695150

A:Note: the translated sequence in Genbank entry HSU12707 (PIDN:AA62663.1) differs f

Proc. Natl. Acad. Sci. U.S.A. 92, 4706-4710, 1995

A:Title: Identification of mutations in the Wiskott-Aldrich syndrome gene and charact

A:Reference number: I38931; MUID:95273432; PMID:7753869

A:Accession: I38931

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-502 <KMA>

A:Cross-references: EMBL:U19927; NID:9854672; PIDN:AAC50140.1; PID:9854673

C:Genetics:

A:Gene: GDB:WAS; IMD2; WASP

A:Cross-references: GDB:120736; OMIM:301000

A:Map position: Xp11.23-Xp11.22

A:Note: defects in this gene may result in Wiskott-Aldrich syndrome

C:Keywords: immunodeficiency

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	0.0328	502	74	24	102	
Percent Similarity:	125.50					
Best Local Similarity:	35.00%					
Query Match:	26.43%					
DB:	2				80	13

US-09-939-293-1 (1-1358) x A55197 (1-502)

```

QY 1291 CCTTCCTTAGTAAATAGGTTT---TTCACTCTGCGCTGACGCTGCTCACAGAC 1235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 ProaPleIethSerSerArgTyrArgGlyLeuProAlaProGlyProSerProAlaasp 224
QY 1234 -----AGTGGGGGACATCAGAGAACAACATACAGAAATACAAAGAAATCGTAC 1184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 LysLysArgSerGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 244
QY 1183 AAACGAGACAGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 Lys-----HsValSerHsValGlyTyrPaspProGlnasn 256
QY 1135 GGGACAAGTACTAATATCTTTTTCAGCAGCTAATTAAGACTGAAGACAGTTAAACAGT 1076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 Gly-----PheAspValAsnAsnLeuAspProAspLeuArgSer 269
QY 1075 TGCTCACTTAAGGACACAAAAGACCTCTCTCTGACCCAGTAGCAAAATGC 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 LeupSerArgAlaGlyLysSerGlu-----AlaGlnLeu 281
QY 1015 TTTGGGTGAGGTTAAATAATGGGTAAAGACAGCTGACAGTGGGGGT----- 965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 ThrAspAlaGlnThrSerLysLeuLysPheLysPheLysPheLysPheLysPhe 301
QY 964 -----GAAATGTAAACAGGGTGCAGTCCCA-----AGG 935
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 AlaValArgGlnLysMetArgArgGlnGlnLeuProLeuProProProProSerArg 321
QY 934 GCTAAGACCGAGTCCGACGCAAGCTGACACACACACACACACACACACACACAG 875
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 GlyGlnAsnGlnLeuProArg---ProProIleValGlyGlnLys----- 336
QY 874 CGTCTCGCCTGATTTGGCCAGGAGGAGCT-----GCCGCC 839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 -----GlyArgSerGlyProLeuProProValProLeuGlyLysLeuPro 351
QY 838 TCTTCTCGTGCACAGACAGTCAATGCCAACCTGGGAGGGTGCATCTGCCCTTT 779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 ProProProThrProArgGlyProProProProGlyArgGlyProProProProPro 371
QY 778 CCCCACTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 ProProAlaIleThrGlyArgSerGlyProLeuProProProProProProProPro 391
QY 718 TCTCTGCTCGACATCAGCCGCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 ropromerProProProProProProProProProProProProSerSerGly----- 407
QY 658 TATCTGCTCTCTCTGACAGCTGGTTTCTGCTTTCCGAGAGACGTGGTACACCTTTCCAC 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 -----AsnGlyProAlaProProProProProProProProProProPro 417
QY 598 CTGACGTTTCCACGCTGAAT---GTGATTCCTGGCGGTATAGAGGCTGATC 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 roAlaLeuValProAla-GlyGlyLeuAlaProGlyGlyGlyArgGlyAlaLeu 434

```

RESULT 2

hypothetical protein BLD1.390 [imported] - Neurospora crassa
 C:Species: Neurospora crassa
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C:Accession: T49385
 R:Schulte, U.; Aign, V.; Hohenseil, J.; Brandt, P.; Farthmann, B.; Holland, R.; Nyekatu
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: T25022

A:Accession: T49385
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-805 <SC>
 A:Cross-references: EMBL:AL355927; GSPDB:GN00116; NCSP:BLD1.390
 A:Experimental source: BAC clone BLD1, strain OR/4A
 A:Gene: NCSP:BLD1.390
 A:Map position: 6
 A:Introns: 54/3; 212/3

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	0.0554	805	49	18	72	
Percent Similarity:	123.00					
Best Local Similarity:	42.41%					
Query Match:	31.01%					
DB:	2				19	6

US-09-939-293-1 (1-1358) x T49385 (1-805)

```

QY 927 ACCAGTCCAGGCGCAAGCTGAGACACAGGAGCATCACAGCTCACAAAGCGCTCG 868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 ThrArgLysSerAlaThrProThrProSerGlnThrGlyThrProGlnProProIleVal 422
QY 867 CCTGATTCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 ThrProIleProProProProAlaValProProValAlaGlyAlaProThrProValProAla 442
QY 807 CTGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 Ala-AlaAlaGlySerLeuAlaAlaLeuProValAlaAlaValAlaValArgSerGly 462
QY 762 GACA-----GGGCGATGTGCT-----CAGGCCCTCAATCTCTCAGCGCAGG 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 462 nThrProThrValGlyProAlaAlaProProProLeuLysLysLysLysLysLysLys 482
QY 723 TAGGCTCTGCTCGGACTCAGCCCGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 rAlaProProThrValProAlaAlaAlaProProProProProProProProProPro 501
QY 663 TCTTCTATCTGCTCTGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 AleuLeuAlaMetLeuArgGlnSerGlyLeuLeuSerGlyThrProAlaAlaProAl 521
QY 603 TCACCTGACGTTTCCACGAGCTGAATGTGAT-----TCTTGGCGGTTATAGAGGCTGGA 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 aProValAlaThrAlaProAlaAlaAlaProAlaAlaAlaAlaAlaAlaAlaAlaAla 538
QY 549 TCTGGCGCAGTTTGTATATGATGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 rLeuLysGlnPheArgGlyProGlnLeuLysAsnArgLeuHisGlnLysPleu 555

```

RESULT 3

hypothetical protein F59A2.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T22976; T23157
 R:Lightning, J.
 submitted to the EMBL Data Library, June 1994
 A:Reference number: Z19645
 A:Accession: T22976
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1133 <MTL>
 A:Cross-references: EMBL:Z234801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F59A2.6
 A:Experimental source: clone F59A2
 R:Burton, J.

N-methyl-D-aspartate receptor chain NMDAR2D-1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A45219
C:Rishi, T.; MoriYoshi, K.; Sugihara, H.; Sakurada, K.; Kadocani, H.; Yokoi, M.; Akazawa
J. Biol. Chem. 268, 2836-2843, 1993
A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor s
A:Reference number: A45219; MUID:9315102; PMID:8428958
A:Accession: A45219
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1356 <ISH>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP:124264)
C:Superfamily: N-methyl-D-aspartate receptor 2D, glutamate receptor homology
F:451-879/Domain: glutamate receptor homology <GRH>

Alignment Scores:	
Pred. No.:	0.571
Score:	111.00
Percent Similarity:	31.378
Best Local Similarity:	27.458
Query Match:	4.608
DB:	1
length:	1356
Matches:	56
Conservative:	8
Mismatches:	73
Indels:	67
Gaps:	9

US-09-939-293-1 (1-1358) x C45219 (1-1356)

QY	915	GCAAGCTGTAGACACAGAGGAGGACATCAGAGCTCACAAGAGGCGTCCGATATGGGACA	856
Db	1004	AlaIleValArgIuInGluInGluProThrIrgIuProProAlaGlyAlaPheProGlyI-PhePro	10233
QY	855	GGGCGAGACCTGCGCGCTCTTCGCTGCACAGACAGTCATGCCAACCTGGGCGAGGTG	796
Db	1024	SerProProAlaProProAlaAlaAlaAlaAlaAlaValIglyProProLeuGlySerIrgu	10433
QY	795	GCA-----TCTGCCCCCTGCTTCCCCACTGAGTGG-----	766
Db	1044	AlaPheGluAspIuSerIuSerProProAlaProSerATgTTPProIArgSerAspProGluSer	10633
QY	765	-----GGAGACAGGCGAGTGTCTCAGGCGCCATCCGTCACGGAGGTAGGCG	718
Db	1064	GluProLeuLeuGlyIleGlyAla-----GlyGlyProSerAlaGly-AlaProThrAl	1081
QY	717	TCTGCTCGAGCTACGACCGGCTCTCTCC-----CTTCTCTCTGTTG	676
Db	1081	aproProProArgIArgAlaAlaProProProGlyAlaTyrIleAspIrguInGluProSerP	1101
QY	675	TTCTGAC-----GGAGCTCTTATCTGTGCTTTCTGCACG	640
Db	1101	oSerAspSerGluAspSerGluSerIeuGlyIleAlaSerIeuGlyIleGluProTr	1121
QY	639	TTGGTTCTG-----CTTTCGGGAGAGCGTGTGCACCT-----	606
Db	1121	pTTPheAlaAspPheProTyrIProTyrAlaGluIuArgIeuGlyProProGlyAlaTgTy	1141
QY	605	-----CTTGCACCTGC	595
Db	1141	rTTPSerValAspIysIeuGlyGlyTTPrIrgAlaGlySerTTPAspTyIrguProProTr	1161
QY	594	AGTTTACCGACT-----GATGTGATTCCTGGCGGTATATAGAGCGCTGATCT	547
Db	1161	gGlyGlyProAlaTTPrHisCysATgHisCysAlaSerIeuGluIeuLeuGluProProIArg	1181
QY	546	GGCGCAATTGATATGCAGCTTCTGGTCGCAATCTCGAAGACCAACGACAGTATCCA	487
Db	1181	gHis-----LeuSerCysSerHisAspGlyIleuAspGlyIleGlyTyr	1194

RESULT
T10627

hypothetical protein T13K14.10 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse ear cress)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
 C:Accession: T10627
 R:Beyan, M.; Pohl, T.; Weizenecker, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Len
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: Z16991
 A:Accession: T10627
 A:Molecule type: DNA
 A:Residues: 1-1396 <BEV>
 A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP-T13K14.10
 A:Experimental source: cultivar Columbia; BAC clone T13K14
 C:Genetics:
 A:Gene: ATSP-T13K14.10
 A:Map position: 4
 A:Initrns: 145/1; 171/2; 196/1; 242/3; 269/3; 298/3; 347/3; 374/3; 414/3; 442/3; 505/
 131/3; 1139/3; 1163/3; 1194/3; 1226/3; 1259/3; 1290/3; 1344/3

Alignment Scores:	
Pred. No.:	1.11
Score:	107.50
Percent Similarity:	36.15%
Best Local Similarity:	22.54%
Query Match:	4.39%
DB:	2
Length:	1396
Matches:	48
Conservative:	29
Mismatches:	71
Indels:	65
Gaps:	6

US-09-939-293-1 (1-1358) x T10627 (1-1396)

QY	203	AAATCAGAGCCTTACCTCCCT	-----	223
Db	1128	LysaspasnProHlsatgdeuValLysleuaspalaProgluLnuasplysAlaIa	1147	223
QY	224	-----AGTACGAAGCATTTGATGAGAGACGAGTCTTTGGTAAACAGATAGC	271	
Db	1148	SerAlaProThrCysSerLysSerValSerGluArgLengluIngluValArgaspHr	1167	
QY	272	ACCTCTACCTTTCTCTCTCAGACCACATATGCGTTGATTAAGAGCTATTACGAATATACT	331	
Db	1168	LyslleLysPheLengLysaLeuLysGlngluThrGluIngluArgSerLwlrParg	1187	
QY	332	AAGCGCTTTATACCTTAACCTCTCTTTTACCGACAATATACAAAGTTTACTTGGGAAATG	391	
Db	1188	LysleuYsThrCysleuLysSerGluTyrrProAspTyThrProleuAlaLyslle	1207	
QY	392	-----AATTACAGAGGAGGAAGATGAAGTGTGCAGGTGATCATAGCA	433	
Db	1208	LeuIngluLysleuLeuSerArgSerAspAlaGlyAspLyslle	1221	
QY	434	GCCAGAGCTGAGATGACTTCAAAACCCAAAGAGTCTTGAACCTGGAAACCACTGGATG	493	
Db	1222	-----SerHlshISgIuGluHlleGluAlaAlaIasnGluVal	1235	
QY	494	ACTGCAGTTGGTCTTTTCAGAGATGGCAGCAGAAAGTGCATATCAAACTGGCGACAGATCAG	553	
Db	1236	ArgSerValaspAlaspGluLeu	1243	
QY	554	GCCCTTATAACCCGCCAGGAATCACTATTCAGCTGGTAAACCTCAGAGTGAAGAGGTGCAC	613	
Db	1244	-----AlaArgPheLeuLeuAspLysThrGluProGluaspGluAlaGlu	1259	
QY	614	CAGCTCTCCCGGAAAGCAGAAACCAAG	640	
Db	1260	LysleuLysLysLysMetGluValThrArgaspGlnleuAlaIaspAlaLeuTyrrGlnLys	1279	
QY	641	---CTGGCAGAAAGCAGATGAAGAAGCTCCGTACAGAAAACACAGAGAGAAACGGAGAG	697	
Db	1280	GlyleuAlaMetAlaArgIleGluAsnleuLysGlyGluTyrgluGlyGluGlu	1299	
QY	698	CGGGCTAGTCCGACAGGAGGCCCTACCTGCGGTAGAGAT	736	
Db	1300	-----GlnSerSerGlnLysaspLysPheGluGluAsn	1310	

RESULT 8


```

Db 2639 -----ProValGlyProCargAspAspPhear 2647
QY 1054 AAAA---AGGACTCTCTCTGACCCAGTAGGCAAAATGCTTTGGGCTGACGTAATA 998
Db 2647 gArGLeuProGInSerSerProAlaProProAlaThrAlaProAlaArgProBr 2667
QY 997 AAAATGGGTAAAGACAGCTGTACAGAGTGGGTGAATTTAAAGAGGGTGCAGAGCCCA 938
Db 2667 oAlaSerSerArgAla-----SerAlaAlaSerSerSe 2678
QY 937 AGGGCTAAGAACCCAGCTCAGCGCAGCCTGAGACCAGAGGAGCAGCTCAGCTCACA 878
Db 2678 rGlySerArGAlaArGArGHisArGArGAlaArGSerLeuAlaArGAlaThrGlnAlaSe 2698
QY 877 AGGCTTCGCGCTGATTTGGCCAGGCGAGGAGCAGCTCCCTCTTTCGTTGGTGCACAGACGT 818
Db 2698 rAlaThrThrGlnGlyTrp-----ArgProProAlaLeu---ProAspThrVa 2713
QY 817 CATGGCAACCCCTGGGCGAGGGTG---GCATCTGCCCTCTTCCGCACTGAGTGGGGAGA 761
Db 2713 lAlaProValThrAspPheAlaArGProProAlaProProAlaProProGlnProAlaPr 2733
QY 760 CAGGCAAGTGTCTCAGGCGCTC-----AATCTCAGCAGGAGTAGGCTCTCTG 713
Db 2733 oHisAlaLeuValSerGly-ValProLeuProLeuGlyProGlnAlaAlaGlyGlnAlaAs 2753
QY 712 CTCCGACACAGCCGCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 653
Db 2753 ePrProAlaLeuProAlaAsProAlaProAlaProAlaThrGly-----T 2769
QY 652 TGCTTCTGCACAGCTTGTTT-----CTGCTTCCGGAGAGAGCTGGTGAC 608
Db 2769 hrValLeuProGlyGlyGlnuAsnArGArGProProLeuThrSerGly-----ProAlaP 2787
QY 607 CTCTTCCACCTGCAGCTTTCACAGCTGAATGTGATTCCTGGGGGTAAAGAGCGCTGATC 548
Db 2787 rOthProProArGVal---ProValGlyGlyProGlnArGArGLeuThrArgPro---- 2804
QY 547 TGGCGCAGTTGATATGACAGCTTCGTCGCCATCTCTGAAACCAACGACGACGTCA---- 492
Db 2805 -----AlaValAlaSerLeuSerLeuSerArGArgGluSerLeuP 2817
QY 491 --TCCAAGTGTTCAGAGTTCAGTCAAGTACTGTTGTTT----- 453
Db 2817 rOserProTrpAspProAlaAspProThrAlaProValLeuGlyArgAsnProAlaGluP 2837
QY 452 --AAGTCATCTCAGCTGTGGCTCTATGATCATGTCAGCACACTTCATCTTCTCTCT 399
Db 2837 rOthSerSer-----SerProAlaGlyProSerProProPro 2849

RESULT 10
G02520
C:Species: human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
C:Accession: G02520
R:McLean, W.H.I., Smith, F.J.D.
submitted to the EMBL Data Library, March 1996
A:Reference number: H01385
A:Accession: G02520
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4574 <NCCL>
A:Cross-references: EMBL:U53204; NID:g1477645; PIDN:AB05427.1; PID:g1477646
C:Genetics:
A:Gene: PLECL1
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S
F:68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

Alignment Scores:
Pred. No.: 1.84 Length: 4574
Score: 105.50 Matches: 38
Percent Similarity: 47.62% Conservative: 32

```

```

Best Local Similarity: 25.85%
Query Match: 4.31%
DB: 2
Gaps: 6
US-09-939-293-1 (1-1358) x G02520 (1-4574)
QY 386 AAAATCAATTCAGAGGAGGAGAAATGATGATGTGGCAGTGATCATAGACCGAGCTGAG 445
Db 2073 ArgLeuGlnAlaGluGlnLysAlaHisAlaPheAlaVal-----GlnGln 2087
QY 446 ATGACTTCAAAACACCAAGACTACTTGAAGCTGGAACCACTGTGATGACTCAGT--- 502
Db 2088 LysGluGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2107
QY 503 GGTCTTTGACAGATGGCAGCAGAGAGCTGCATATCAAACTGGCGCAGATAGGCTTATA 562
Db 2108 GlyGlnAlaGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 2125
QY 563 ACCGAGGAAATCACACTTCACTGCTGAGTGAATGCAAGCTGGAGAGAGGTGCACAGCTCC 622
Db 2126 GlnAlaGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 2145
QY 623 CGGAAAGCAGAAACCAAGCTG-----GCAGAAACAGATAGAGAGCTC 667
Db 2146 GlnSerAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2165
QY 668 CGTCAGAAAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 727
Db 2166 ArgLysGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2184
QY 728 CGTGAGGATTGAGGGGCTGAGCAGACTGCCCTGTCCTCCCTCAGTGGGAAAGCAGGG 787
Db 2185 ArgGln-----LysGlnAla 2189
QY 788 GCAGATGCCACCCCTGCCAGG 808
Db 2190 AlaAspAlaGlnMerGluLys 2196

RESULT 11
G90968
EspF-like protein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: G90968
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gatawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90968
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA86138.1; PID:g13362183; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS2715

Alignment Scores:
Pred. No.: 1.54 Length: 337
Score: 105.00 Matches: 66
Percent Similarity: 31.70% Conservative: 31
Best Local Similarity: 21.57% Mismatches: 94
Query Match: 4.35% Indels: 115
DB: 2 Gaps: 13
US-09-939-293-1 (1-1358) x G90968 (1-337)
QY 1323 CTGTTGACAGCACAGAGTAACACTCTGCACCCCTTCTTGTAGTAAATAGGTTTTTCA 1264
Db 55 LeupheArgAlaProAsnAlaHisSerAlaSerPe----- 66
QY 1263 CTGCTTGGCTCAGCTGTCTCTACACAGACAGTGGGGCGATCAGAGAACATCAGAAA 1204

```


A:Accession: T10927
A>Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-517 <PAR>
A:Cross-references: EMBL:AL031231; NID:e1315070; PID:e1315088
A:Experimental source: strain A3(2)
C:Genetics: A:Gene: 3C3.18c

Alignment Scores:

Alignm. NO.:	1.77	Length:	517
Pred. Score:	104.50	Matches:	44
Percent Similarity:	37.50%	Conservative:	19
Best Local Similarity:	26.19%	Mismatches:	74
Query Match:	4.27%	Indels:	33
	2	Gaps:	5

Ds: -09-939-293-1 (1-1358) x T10927 (1-517)

OY 424 GATCATAGACACCAGCTGAGATCACTTCAAACCCAGAGTAGTAACTGAAGCTGAAC 483
Db 185 GTTTTGTTCATAAGTGGTCTTTCAGAGATGCACAGAAGCTGCATTAACAACGCG 204
OY 484 CACTTGATGATGACTGAGTGTCTTTCAGAGATGCACAGAAGCTGCATTAACAACGCG 543
Db 205 HTSPHGLTASPTHTASPTTP----- 211
OY 544 CGCAGATGACGCCCTTATATACGCCGAGGATACATTCAGCTGTGTAAGTGAAGTGA 603
Db 212 ---LysProGlyGluTyrIleSerValAlaProSerHisGlyAlaSerGlyIleGly 230
OY 604 AGAGTGCACCAAGCTCTCCCGAAGACAGAAAACCAAGCTGCAGAGAACAGATAGAAGA 663
Db 231 SerGlyThrProAlaValAlaIleGlyIleAspSerGlySerArgValGlyAlaGln 250
OY 664 GCTCCGTGAGAAAAACACAGAGAGAGAGGGGAGAGCGGCTG----- 704
Db 251 ---ProHisProTyrSerGlyAlaSerGluGlyAlaSerProGlyLeuSerSerAlaSer 269
OY 705 AGTCGAGCAGAGAGCCCTACCTGCGTGAAGATTGAGGCTGAGAGCACCTGCCCTG-TCT 763
Db 270 AlaAspSerArgArgProValAlaMetasPileAspGlyLeuThrThrLeuProGlnArg 289
OY 764 CCCCACTAGTGGGAAAGCAGGGGCGAGATCCACCCCTGCCACG----- 808
Db 290 ProGlyProThrProGlnProGlyProAsnThrProLeuProArgSerAspAlaProPro 309
OY 809 -----GTTGGCATGACTGTCTGTGCACCGAGAAAGCGCGCAGGTCTCTGCCCT 856
Db 310 GlyGlnProValValGlyProProValLeuGlnAsnSerGlyGlyGlyThrAsnAsnPro 329
OY 857 GGCCAATCAGGAGAGACGCTTTG 880
Db 330 Ala---SerGlyArgMetProVal 336

RESULT 14
E75383
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75383
R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.D.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Ma-
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: AF5250; MUID:20036896; PMID:10567266
A:Accession: E75383
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-581 <WHIT>
A:Cross-references: GB:AEO01998; GB:AEO00513; NID:g6459302; PIDN:AF1109.1; PID:g6459303
A:Experimental source: strain R1

Alignment Scores:			
Pred. No.:	1.79	Length:	581
Score:	104.50	Matches:	60
Percent Similarity:	41.42%	Conservative:	39
Best Local Similarity:	25.10%	Mismatches:	75
Query Match:	4.27%	Indels:	65
DB:	2	Gaps:	12
US-09-939-293-1 (1-1358) x E75383 (1-581)			
QY	197 GCACAGAAATCAGACGCTCATCTCCCTTAGTACGAGACATGTGATGAGAGACAGCTGCT	256	
DB	95 AlaleuthrsergluValAlthrAspLeuGluaNGlu-----LeuAlaAspValLeuAla	112	
QY	257 TTGGTACACAGATAGCACCTCTACCTTCTCTCT-----CAGACCAATAT	301	
DB	113 ArgValAlthrAlaAlthrArgAlaLeuArgGluThrGluAlaGlnAlleArgValThrArg	132	
QY	302 GCGTTGATGAGTGAAGTATTACTGAATATPACAAAGCGGTTTAACTTACTTCTCTTAC	361	
DB	133 SerGlnValGluAlaLeuLysValaAspAlaArgAlaVal-----MetLysAlaLeuThr	150	
QY	362 CGA-----CAATATACAGTTACTTGGGAAATGAAATTC-----	397	
DB	151 ArgAlaArgAsnThrGlnTyrMetAlaArgLeuLeuSerGlnSerAsnSerlleSerAspMet	170	
QY	398 -----GAGCAGGAAGATGAAGCTGCGCAGGTG	424	
DB	171 LeuIleArgLeuAspTyrAlaAsnMetAlaGlyGlnArgAsnValGluValMetGluGln	190	
QY	425 ATCATAGAGGCGACAGCTGTGATGACTGTCAAAAC-----	460	
DB	191 LeuArgGlyAlaAlaAlaGluLeuThrThrGlnGlnLeuArgGlnThrGlnSerAsp	210	
QY	461 -----CAAGAGTACTTGAACTG-----GAAACCACT	487	
DB	211 AlaLeuArgGlyLeuGlnGlnGlnGlnGlnInThrLysLeuAlaGluLeuArgAspArgArg	230	
QY	488 TGGATGACTGCATGTGCTCTTTACAGAGTGGCAGCAGAACT-----	529	
DB	231 ThrArgGlnAlaAspAlaLeuAlaGluLeuGlnArgSerAlaGlnGlyGlnAlaVal	250	
QY	530 GCATATCAAACTGGCGGAGATCAAGGCTCTATATACCGCAGGAATACATTCAGCTGGTG	589	
DB	251 AlaValArgThrGlnAlaGlnAlaGlnAla-----LeuThrAlaGlnThrIleAspSerLeuVal	269	
QY	590 -----AAACTGCAGGTGGAAGAGGTGCACACAGCTCCCGG	625	
DB	270 GlyAsnValValaArgGlyGluArgThrArgLeuGlnGlnGlnArgArgArgArgLeuGlnGlu	289	
QY	626 AAAGCAGAAACCAAGTCGGCAGACGA-----CAGATGAAGAGCTCCGTACAGAAAACAG	682	
DB	290 GluArgArgArgArgGluAlaGlnAlaArgArgIleArgGluAlaGlnGlnGlnArgArg	309	
QY	683 GAGCAGAGGGAGAG-----CGGCGTGAAGTCGGAGCAGACAGAGCCTACCTGCGTGA	733	
DB	310 LysGluAlaGlnAlaArgLeuAlaArgIleArgValaGlnGlnGlnArgLysAlaArgGlu	328	

RESULT 15

A56136

tagged protein precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change 11-Jan-2000

C:Accession: A56136

R:Rindsdal, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.

Cell 80, 909-917, 1995

Article: tagged: a mammalian ligand that activates Notch1.

A:Reference number: A56136; M0ID:952181442; PMID:7697721

A:Accession: A56136

Db 313 ThrTrpThrAlaProProThrglyGlySerSerLysAspProSerSerLeuLeu 332
OY 461 GGTGTTTG 453
Db 333 GlyThrLeu 335

RESULT 4
US-08-882-046-5
Sequence 5, Application US/08882046
Patent No. 6136952

GENERAL INFORMATION:
APPLICANT: Li, Linheng
APPLICANT: Hood, Leroy
APPLICANT: Krantz, Ian D.
APPLICANT: Splinter, Nancy B.
TITLE OF INVENTION: Human Jagged Polypeptide, Encoding
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
City: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,046
FILING DATE: 25-JUN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-0W 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1219 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-882-046-5

Alignment Scores:
Pred. No.: 0.0844 Length: 1219
Score: 104.50 Matches: 71
Percent Similarity: 29.67% Conservative: 37
Best Local Similarity: 19.51% Mismatches: 131
Query Match: 4.27% Indels: 125
Gaps: 19

US-09-939-293-1 (1-1358) x US-08-882-046-5 (1-1219)

OY 481 AACCACTGGATGACTGAGTGGCT----- 507
Db 666 AsnAsnIleAsnAspCysSerGlnAsnProCysHisTyrGlyGlyThrCysArgAspLeu 685
OY 508 TTCAGAGATGGCAGCAGAAAGCTGCATATCAACTGGCGCAGATCGAGCGCTCATTAACCGC 567
Db 686 ValAsnAspMetTyrCysAspCysLysAsnGlyTyrIlePlySgLYLysThrCysHisSerArg 705
OY 568 CAGGAATCACATTCAGCTGTGAAGCTGCAGAGTGGAGAGGTGGACCAAGCTCTCCCGGAA 627
Db 706 AspSer-----GlnCysAspGluAlaIleThrCysAsnAsnGlyGlyThrCysTyrAspGlu 723
OY 628 -----AGCAGAAACCAAGCT 642

Db 724 ValAspThrPheLysCysMetCysProGlyGlyTyrPglGlyThrThrCysAsnIleAla 743
OY 643 GGCAGAGCAGACAGATPAGAAGAGCTCCGTCAGAAAACACAGAGAGGAGGAGCGCGC 702
Db 744 ArgAsnSerSerCysLeuProAsnProCysHisAsnGlyGly----- 757
OY 703 TGAATCGAGCAGAGAGGCGCTACCTGAGATTTAGAGGCGCTGAGCAGACTGGCGCTGTC 762
Db 758 -----ThrCysValVal 761
OY 763 TCCCCACTCAGTGGGGAAGCAGGCGCAGATCCACCTGCGCAGGCTGGCATGACTGT 822
Db 762 -----AsnGlyAspSer---PheThrCysValCysLysGlyGlyTyrPglGlyPro 777
OY 823 CTGTGACCCGAGAGAGAGCGCGCAGTCTGCGCCGCGCA----- 861
Db 778 IleCysThrGlnAsnThrAsnAspCysSerProHisProCysTyrAsnSerGlyThrCys 797
OY 862 ATCAGCGCAGACGCGCTTGTGAGCTGTGAGTGCCTCTGCTGAGCTGAGCGCTGGCGA 921
Db 798 ValAspGlyAspAsnTyrTyrArgCysGlyCys-----AlaProGlyPheAlaGly 814
OY 922 CCF-----GGTTTACGCGCTGGGCACTGCACCTGCTTTTAACATTTCACCC 969
Db 815 ProAspCysArgIleAsnIleAsnGlyCysGlnSerSerProCys---AlaPheGlyAla 833
OY 970 ACTCTG-----TACAGCTGCTCTTACCATTCTTTTATCTACACAC 1011
Db 834 ThrCysValAspGluIleAsnGlyTyrGlnCysIleCysPro----- 847
OY 1012 CAAGACATTTTGCCTACCTGGGTCAAGAGAGAGTCCCTTTTTCATGCGCTTAAGTTC 1071
Db 848 -----ProGlyHisSerGlyAlaLysCysHisGlyValSerGly 860
OY 1072 AGCAACGTTTAAGCTGTTCAGCTGTATTAGCTGTCAAAATGATTTAGTACTGT 1131
Db 861 ArgSerCysIleThrMetGlyArgValIleLeuAspGlyAlaLysTyrPAspAspAspCys 880
OY 1132 TCCCTCTGT-----TGGATGCGCAGTGT 1155
Db 881 AsnThrCysGlnCysLeuAsnGlyArgValAlaCysSerLysValTyr-----Cys 897
OY 1156 GGC-----AGGGGAGGGGAACCTGTCCA----- 1179
Db 898 GlyProArgProCysArgLeuGlnIleGlyGlyHisGlyCysProAsnGlyGlnSerCys 917
OY 1180 GTTGTGACGATTTCTTTGATGATTTCTGATGATGCTTCTGATGATGCGCCCACTGT--- 1236
Db 918 IleProValIleuAspAspGlnCysPheValArgProCysThrGlyAlaGlyGlyCysArg 937
OY 1237 -----CCTGTGAGCAGACCTGAGCGCAAGAGTGAAGAAACCTATTACTAC 1281
Db 938 SerSerSerLeuGlnProValIleThrLysCysThrSerAsp-----SerTyrTyr--- 954
OY 1282 TAAAGAAAGGGGTCCAGAGCTTTTAACTGCTGCTCAACAGACTTAACATCAACAGGA 1341
Db 955 ---GlnAspAsnCysAlaAsnIleThrPheThrPheAsnGlyGlnMetSerProGly 973
OY 1342 CTTAACACAGAA 1353
Db 974 LeuThrThrGlu 977

RESULT 5
US-09-308-022-6
Sequence 6, Application US/09308022
Patent No. 6291654

GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORTING, RAASCH & GEBHARDT, P.A.

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,035A
FILING DATE: 29-SEPT-97
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/231,193
FILING DATE: 20-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/052,449
FILING DATE: 20-APR-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9383E
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-940-035A-54

Alignment Scores:
Pred. No.: 0.375 Length: 1214
Score: 98.00 Matches: 65
Percent Similarity: 33.58% Conservative: 25
Best Local Similarity: 24.25% Mismatches: 92
Query Match: 4.06% Indels: 86
DB: 4 Gaps: 12

US-09-939-293-1 (1-1358) x US-08-940-035A-54 (1-1214)

QY 1009 TGTGAGGTAAATAATGGCTAAGACAGCTGT----- 977
DB 913 CysProthPrProArGserGlyProserProCysLeuProThProAsrProProProGlu 932
QY 976 -----ACAGAGTGGGGTGAATGTTAACAGGGTGCACAGTCCCAAGGGCTAAGAAC 926
DB 933 ProserProThGlyTrpLy----- 939
QY 925 CAGGTTCAGAGCGCAGCTGACACAGAGGAGGACACTACAGCTCACAAGGCGTCCGCC 866
DB 940 -----ProProAsp-----GlyGlyArgAlaAlaLeuValArgArg-AlaPr 953
QY 865 TGAATGGCCAGGAGAGAGCTGCCGCTCTTCTCGGTGCACAGAGTATGCCAACCCCT 806
DB 953 ogLnProProGlyArg-----ProProThrProGlyProPro-----LeuSerAspVa 969
QY 805 GGGCAGGAGGAGTGTGCGCCCTGCTTCCCACTAGTGGGAG-----AAGGGCAAT 752
DB 966 lSerArgValSerArgArgProAlaTrpGluAlaArgTrp-ProValArgThnGlyHnSc 989
QY 751 GTGCTCAGGCGCCCAAT----- 735
DB 989 ysgLyArgHnSerAlaSerGluArgProLeuSerProAlaArgCysHnTySerS 1009
QY 734 -----CTTCAGCCAGAGGAGCTCTCGTCGACACTCAGCCGCTCCCTCTCTCTG 680
DB 1009 erPhseProArgAlaAspArgSerGlyArgProPhelauProLeuPhelauProGluProG 1029
QY 679 TGTTTTCAGAGGAGCTCTTCA-----TCTGTCCTTC 647
DB 1029 luleuGlnuSpLeuProLeuLeuGlyProGluGlnLeuAlaArgArgGluAlaLeuLeu 1049
QY 646 TCCACAGCTTGCTTCTGCTTCCGGGAGAGTGTGACACCTCTTCCACCTGCAGTTTCA- 588

DB 1049 snAlaAlaTrpAlaArgGlySerArgProSerHnAlaSerLeuProSerSerValAlag 1069
QY 587 -----CAGCTGAATGTATTCCTGGCGGTTATAGAGCC 554
DB 1069 lualAphAlaArgProSerSerLeuProAlaGlyCysThrGlyProAlaCysAlaArgP 1089
QY 553 CTGAT-----CTGGCAGATTGATMGAGCTTCTGTCGCAT 515
DB 1089 roAspGlyHnSerAlaCysArgArgLeuAlaGln---AlaGlnSerMetCysLeuProI 1108
QY 514 CTGTGAAGAGCCAGCTGCATGCATCCAGTGGTTCCAGCTTCAAGT----- 468
DB 1108 lEtyArgGluAlaCysGlnGluGlnAlaGlyAlaArgAlaTrpGlnHnArgS 1128
QY 467 --ACTCTTGAGTTTGAAGTATCTCAGCTGTGCTGCTTAAGTACCTGCCACACTTC 410
DB 1128 lHnIsValCysLeuHnIsAlaHnIsAlaHnIsLeuProLeuCysTrpGlyAlaValCysProH 1148
QY 409 ATCTTCCTCTGATTC 390
DB 1148 lSLeuProCysAspSer 1154
RESULT 11
US-08-935-105A-54
Sequence 54, Application US/08935105A
Patent No. 637660
GENERAL INFORMATION:
APPLICANT: Daggett, Lorrie P.
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehrman White & McCauliffe
STREET: 4250 Executive Square, 7th Floor
CITY: La Jolla
STATE: CA
COUNTRY: U.S.A.
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/935,105A
FILING DATE: 29-SEPT-97
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/231,193
FILING DATE: 20-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/052,449
FILING DATE: 20-APR-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9383D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0062
TELEFAX: 619-238-0999
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-935-105A-54
Alignment Scores:
Pred. No.: 0.375 Length: 1214
Score: 98.00 Matches: 65

Percent Similarity: 33.58% Conservative: 25
 Best Local Similarity: 24.25% Mismatches: 92
 Query Match: 4.06% Indels: 86
 DB: 4 Gaps: 12

US-09-939-293-1 (1-1358) x US-08-935-105A-54 (1-1214)

```

QY 1009 TGTGAGGTAAATAATGGTAAGACAGCTGT----- 977
    |||  |||  |||  |||
Db 913 CysProthPrOArGSeGgLyProSerProCysLeuProthPrOArSPProProGlu 932
QY 976 -----ACAGAGTGGGGTGAATGTTAAACAGGTCAGTCCCAAGGGCTAAGAAC 926
    |||  |||  |||  |||
Db 933 ProSerProthPrOArGlyTPrGly----- 939
QY 925 CAGGTCCAGGCCAAGCTGAGACCAAGAGGACACTCAGACTCAAAAGGCTGTGCC 866
    |||  |||  |||  |||
Db 940 -----ProProASP-----GlyGlyArGAlaAlaLeuValArGArG-AlAr 953
QY 865 TGATTTGGCAGGAGGAGACCTGCGCCTCTTCGCTGCACAGACAGTCAACCCCT 806
    |||  |||  |||  |||
Db 953 OGlnProProGlyArG-----ProProthPrOArGlyProPro-----LeuSerArPyA 969
QY 805 GGGCAGGTTGGCATCTGCCCCCTGCTTCCCACTGAGTGGGAG-----ACAGGGCAGT 752
    |||  |||  |||  |||
Db 969 lSerArGValSerArGArGProAlArTrPrGluAlArGTrP-ProValArGThrGlyHisc 989
QY 751 GTGCTCAGCCCTCAAT----- 735
    |||  |||  |||  |||
Db 989 ySGlyArGHisLeuSerAlaSerGluArGProLeuSerProAlArGcYshisTyrSers 1009
QY 734 -----CCACAGCAGGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 680
    |||  |||  |||  |||
Db 1009 eRPrOArGAlArSPArGSerGlyArGProPrOArGLeuPrOArGProProG 1029
QY 679 TGTTTTTCGACGAGCTCTCTTA-----TCTGTGCTTC 647
    |||  |||  |||  |||
Db 1029 luleGluAlArSPLeuPrOArGLeuGlyProGluGlnLeuAlArGAlArGluAlaLeuAla 1049
QY 646 TGCCAGCTTGCTTCTGCTTCCGGAGAGCTGGTGCACCTTCCACCTGCACTTCA- 588
    |||  |||  |||  |||
Db 1049 snAlaAlArTrPrAlArGlySerArGProSerHisAlaSerLeuPrOArSerArValAlaG 1069
QY 587 -----CCAGCTGAATGTGATTCGCGGGTTATAGAGGC 554
    |||  |||  |||  |||
Db 1069 lualArPhAlArGProSerSerLeuProAlArGlyCysThGlyProAlArGysAlaArGPr 1089
QY 553 CTGAT-----CTGCGCAGTTTGAATAGCAGCTTGTGCTGCAT 515
    |||  |||  |||  |||
Db 1089 roArSPGlyHisSerAlaCysArGArGAlaGln-----AlaGlnSerMetCysLeuProI 1108
QY 514 CTCGAAAGACCACTGACAGTCAACCAAGTGTTCAGCTCAAGT----- 468
    |||  |||  |||  |||
Db 1108 leryArGAlArGAlaCysGlnGlyGlnAlaGlyAlArProAlArTrPrGlnHisArG 1128
QY 467 --ACCTGCTGTTTGAAGTCACTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 410
    |||  |||  |||  |||
Db 1128 lnhIsValCysLeuHisAlaHisAlaHisLeuProLeuCysTrPrGlyAlaValCysProH 1148
QY 409 ATCTTCTCTCTGCAATCA 390
    |||  |||  |||  |||
Db 1148 lSleuProProCysAspSer 1154

```

RESULT 12

US-09-648-797-54

; Sequence 54, Application US/09648797

; Patent NO. 6469142

; GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.

Ellis, Steven B.

Liaw, Chen W.

; Lu, Chin-Chun

; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR

SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR

NUMBER OF SEQUENCES: 63

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McCaulliffe

STREET: 4250 Executive Square, 7th Floor

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/648,797

FILING DATE: 28-Aug-2000

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/940,086A

FILING DATE: 29-SEPT-97

APPLICATION NUMBER: US 08/231,193

FILING DATE: 20-APR-1994

APPLICATION NUMBER: US 08/052,449

FILING DATE: 20-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24735-9383C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 450-8400

TELEFAX: (619) 450-8499

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 1214 amino acids

TYPE: amino acid

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 54:

US-09-648-797-54

US-09-939-293-1 (1-1358) x US-09-648-797-54 (1-1214)

```

QY 1009 TGTGAGGTAAATAATGGTAAGACAGCTGT----- 977
    |||  |||  |||  |||
Db 913 CysProthPrOArGSeGgLyProSerProCysLeuProthPrOArSPProProGlu 932
QY 976 -----ACAGAGTGGGGTGAATGTTAAACAGGTCAGTCCCAAGGGCTAAGAAC 926
    |||  |||  |||  |||
Db 933 ProSerProthPrOArGlyTPrGly----- 939
QY 925 CAGGTCCAGGCCAAGCTGAGACCAAGAGGACACTCAGACTCAAAAGGCTGTGCC 866
    |||  |||  |||  |||
Db 940 -----ProProASP-----GlyGlyArGAlaAlaLeuValArGArG-AlAr 953
QY 865 TGATTTGGCAGGAGGAGACCTGCGCCTCTTCGCTGCACAGACAGTCAACCCCT 806
    |||  |||  |||  |||
Db 953 OGlnProProGlyArG-----ProProthPrOArGlyProPro-----LeuSerArPyA 969
QY 805 GGGCAGGTTGGCATCTGCCCCCTGCTTCCCACTGAGTGGGAG-----ACAGGGCAGT 752
    |||  |||  |||  |||
Db 969 lSerArGValSerArGArGProAlArTrPrGluAlArGTrP-ProValArGThrGlyHisc 989
QY 751 GTGCTCAGCCCTCAAT----- 735
    |||  |||  |||  |||
Db 989 ySGlyArGHisLeuSerAlaSerGluArGProLeuSerProAlArGcYshisTyrSers 1009

```

Alignment Scores:

Pred. No.: 0.375

Score: 98.00

Percent Similarity: 33.58%

Best Local Similarity: 24.25%

Query Match: 4.06%

DB: 4

Length: 1214

Matches: 65

Conservative: 25

Mismatches: 92

Indels: 86

Gaps: 12

[illegible]

RESULT 13

US-08-231-193A--50
Sequence 50, Application US/08231193A
Patent No. 5849895
GENERAL INFORMATION:
APPLICANT: Daggett, Lorrie P.
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen W.
APPLICANT: Lu, Chin-Chun
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: U.S.A.
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,193A
FILING DATE: 20-APR-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/052,459
FILING DATE: 20-APR-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9383
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
INFORMATION FOR SEQ ID NO: 50:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1219 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-231-193A-50

```

Alignment Scores:

Pred. No.:	0.376	Length:	1219
Score:	98.00	Matches:	65
Percent Similarity:	33.58%	Conservative:	25
Best local Similarity:	24.25%	Mismatches:	92
Query Match:	4.06%	Indels:	86
DB:	2	Gaps:	12

US-09-939-293-1 (1-1358) x US-08-231-193A-50 (1-1219)

QY 1009 TGTGAGGTAATAATGGGTAAGAGCAGCTGT----- 977
 ||| :: ||| |||

db 918 CysProThrProArgSerGlyProSerProCysLeuProThrProAspProProGlu 93/

[illegible]

925 CAGGTCAGCGCAAGCCTGAGACCAACAGAGAGGCCTCACAAGCTCAGAAAGCGCTTCGCC 866

```

Db 945 -----PROProMSP-----GlyGlyIArgAlaAlaLeuValArgArg-AlaPr 956
      |||::: ||||| ::: ||||| ::: |||

```

QY 865 TGATTGGCCAGGCGACCTGCGCCTCTCTCTGCTGACACAGACAGATCATGCGCAACCT 808
... ..

Db 958 oGlnProProGlyArg-----ProProThrProGlyProPro-----LeuSerAspVal 914

[illegible]

0x 751 GTGCTAGGCCCTCAAT----- 735

Db 994 ysglyarghislleuseralasergluargproleuserproalargcyshtyrser 101

oy 734 -----CCTCAGCAGTAGGCTCTCTGCTCCGACTCAGCCCGCTCTCCCTTCCTCTG 680

Db 1014 erPheProArgAlaaspArgSerGlyArgProPheLeuProLeuPheProGluProProG 1033

QY 679 TGTTCGACGAGACTCTTCTA-----TCTGTGCTTC 647
||| ||||| ::::: |||

[illegible]

```

1054 snA|AA|ATrNA|aArAG|vSerArArProSerH|SA|aSerLeuProSerSerVal|Alag 1072
      |||||      |||      ::      |||      |||||      ::|||:::
Db

```

QY 587 -----CCAGCTGATGTGATTCTCGCGGTATAGAGGC 554

Db 1074 IuAlaPheAlaArgProSerSerLeuProAlaGlyCysThrGlyProAlaCysAlaArgP 1099

0Y 553 CTGAT-----CTGCCGACGTTGATATGCACGTTCTGCTGCCAT 511

DB 1094 roaspgylhisseraIacysargargLeuAlaGln---AlaGlnserIetCysLeuProI III

Case	Number of iterations	Number of iterations with gradient	Number of iterations with Hessian	Number of iterations with Hessian and gradient
1112	10	10	10	10
1113	10	10	10	10
1114	10	10	10	10
1115	10	10	10	10
1116	10	10	10	10
1117	10	10	10	10
1118	10	10	10	10
1119	10	10	10	10
1120	10	10	10	10
1121	10	10	10	10
1122	10	10	10	10
1123	10	10	10	10
1124	10	10	10	10
1125	10	10	10	10
1126	10	10	10	10
1127	10	10	10	10
1128	10	10	10	10
1129	10	10	10	10
1130	10	10	10	10
1131	10	10	10	10
1132	10	10	10	10
1133	10	10	10	10
1134	10	10	10	10
1135	10	10	10	10
1136	10	10	10	10
1137	10	10	10	10
1138	10	10	10	10
1139	10	10	10	10
1140	10	10	10	10
1141	10	10	10	10
1142	10	10	10	10
1143	10	10	10	10
1144	10	10	10	10
1145	10	10	10	10
1146	10	10	10	10
1147	10	10	10	10
1148	10	10	10	10
1149	10	10	10	10
1150	10	10	10	10
1151	10	10	10	10
1152	10	10	10	10
1153	10	10	10	10
1154	10	10	10	10
1155	10	10	10	10
1156	10	10	10	10
1157	10	10	10	10
1158	10	10	10	10
1159	10	10	10	10
1160	10	10	10	10
1161	10	10	10	10
1162	10	10	10	10
1163	10	10	10	10
1164	10	10	10	10
1165	10	10	10	10
1166	10	10	10	10
1167	10	10	10	10
1168	10	10	10	10
1169	10	10	10	10
1170	10	10	10	10
1171	10	10	10	10
1172	10	10	10	10
1173	10	10	10	10
1174	10	10	10	10
1175	10	10	10	10
1176	10	10	10	10
1177	10	10	10	10
1178	10	10	10	10
1179	10	10	10	10
1180	10	10	10	10
1181	10	10	10	10
1182	10	10	10	10
1183	10	10	10	10
1184	10	10	10	10
1185	10	10	10	10
1186	10	10	10	10
1187	10	10	10	10
1188	10	10	10	10
1189	10	10	10	10
1190	10	10	10	10
1191	10	10	10	10
1192	10	10	10	10
1193	10	10		

467 --ACTCTGGTGTGTTGAAGTCATCTCAGCCTGGCTCTATGATCACTGCCACACTTC 410

```

Db 1133 INHISValCysLeuNHISAlaNHISAlaNHISLeuProLeuCysTyrGlyAlaValCysProN 115
      ::: :::      :::  |||  |||

```

QY 409 ATCTTCCTCCTGATTC 390

Db 1153 isLeuProProCysaspSer 1159

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: February 20, 2003, 15:01:24 ; Search time 16.5 Seconds

(without alignments)
4205.500 Million cell updates/sec

Title: US-09-939-293-1

Perfect score: 2446

Sequence: 1 ggcgtccgcgcgtgcacaa.....ggacttacacacagaataaaa 1358

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 280518

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```
-MODEL=frame_plus.n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09939293/runat_20022003_111513_15000/app-query.fasta.1.1543
-DB=Published_Applications_AA -QFM=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09939293.eccn1_1_11@runat_20022003_111513_15000
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAPP -LARGEDEERT -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Published Applications_AA.*

```
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PC1US_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1138	46.5	227	10	US-09-925-297-591
2	1016	41.5	237	10	US-09-798-116-2
3	1016	41.5	237	10	US-09-798-116-4
4	1004	41.0	202	10	US-09-798-116-7

5	869	35.5	177	10	US-09-798-116-8	Sequence 8, Appl
6	773	31.6	177	10	US-09-798-116-6	Sequence 9, Appl
7	333	13.6	84	10	US-09-798-116-9	Sequence 6, Appl
8	190	7.8	40	10	US-09-939-293-2	Sequence 2, Appl
9	186	7.6	39	10	US-09-939-293-8	Sequence 8, Appl
10	164	6.7	35	10	US-09-939-293-11	Sequence 11, Appl
11	157	6.4	32	10	US-09-798-116-24	Sequence 10, Appl
12	148.5	6.1	73	10	US-09-798-116-10	Sequence 14, Appl
13	141	5.8	30	10	US-09-939-293-7	Sequence 7, Appl
14	137	5.6	29	10	US-09-798-116-25	Sequence 25, Appl
15	111	4.5	401	9	US-09-764-868-701	Sequence 701, App
16	107	4.4	1509	10	US-09-901-940-2	Sequence 2, Appl
17	103.5	4.2	317	10	US-09-789-561-171	Sequence 17, App
18	98	4.1	1214	9	US-09-945-901-54	Sequence 54, Appl
19	98	4.1	1214	9	US-10-007-747-54	Sequence 54, Appl
20	98	4.1	1214	9	US-10-038-937-54	Sequence 50, Appl
21	98	4.1	1219	9	US-09-945-901-50	Sequence 50, Appl
22	98	4.1	1219	9	US-10-007-747-50	Sequence 50, Appl
23	98	4.1	1219	9	US-10-038-937-50	Sequence 48, Appl
24	98	4.1	1231	9	US-09-945-901-48	Sequence 48, Appl
25	98	4.1	1231	9	US-10-007-747-48	Sequence 48, Appl
26	98	4.1	1231	9	US-10-038-937-48	Sequence 48, Appl
27	98	4.1	1236	9	US-09-945-901-6	Sequence 6, Appl
28	98	4.1	1236	9	US-10-007-747-6	Sequence 6, Appl
29	98	4.1	1236	9	US-10-038-937-6	Sequence 6, Appl
30	98	4.1	1239	9	US-09-945-901-52	Sequence 52, Appl
31	98	4.1	1239	9	US-10-007-747-52	Sequence 52, Appl
32	98	4.1	1239	9	US-10-038-937-52	Sequence 52, Appl
33	98	4.1	1244	9	US-09-945-901-46	Sequence 46, Appl
34	98	4.1	1244	9	US-10-007-747-46	Sequence 46, Appl
35	98	4.1	1244	9	US-10-038-937-46	Sequence 46, Appl
36	94.5	3.9	503	9	US-10-078-547-2	Sequence 2, Appl
37	94	3.8	218	10	US-09-789-561-193	Sequence 193, App
38	93.5	3.9	738	10	US-09-978-979-6	Sequence 6, Appl
39	93.5	3.8	2139	10	US-09-727-384-6	Sequence 6, Appl
40	93	3.9	200	10	US-09-925-297-700	Sequence 700, App
41	93	3.8	300	10	US-09-919-603-7	Sequence 7, Appl
42	92.5	3.8	826	9	US-09-884-988-47	Sequence 47, Appl
43	92	3.8	663	9	US-10-080-960-14	Sequence 14, Appl
44	92	3.8	1617	10	US-09-784-358-16	Sequence 16, Appl
45	92	3.8	1691	10	US-09-784-358-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-925-297-591
; Sequence 591, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 591
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-591
Alignment Scores: 1.26e-92 Length: 227
Pred. No.: 227
```

Score:	1138.00	Matches:	226
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	46.52%	Indels:	0
DB:	10	Gaps:	0

US-09-939-293-1 (1-1358) x US-09-925-297-591 (1-227)

```

OY 59 TCATTTTCAGGTAACAGACAGTGTGTTGTGTCCTGTGTTGGCTAACTTTAAGAGGG 118
    |||||||
Db 2 SerPhepheaYgYfArGInLncysLeucysVAIProvaIvalAIAasnPhelcysysArG 21
    |||||||
OY 119 TGTTCCTCAGATATGATTAAGACATGACCAAAATGTGACAGATGGCTTTGGATGTAAC 178
    |||||||
Db 22 CysPheSerGIuLeuIIleArGProtrPhIsylStrValThrlleGlyPhelcylValThr 41
    |||||||
OY 179 CTGTGTGGGGTTCCTATTGACACAGAAATCAGAGCCTCATTCCTTAAAGTAGAGACATTG 238
    |||||||
Db 42 LeucysAlaValProIIleAlaGInLysSerGIuProHIsSerLeuSerGIuAlaLeu 61
    |||||||
OY 239 AFGAGAGAGCGAGCTGTCCTTTGGTAACAGATAGACCTTACCTTCTCTCTCAGACCA 298
    |||||||
Db 62 MetAlaGArgAlaValSerLeuValThrAspSerTrnSerThrPheLeuSerGIuTrnThr 81
    |||||||
OY 299 TATGGCTTGAATGAAGCTATTACTGAAATATACTAAGAGCTGTTTATACCTTACTCTCT 358
    |||||||
Db 82 TyralaLeuIIleGluAlaIIleThrcGIuTrnThrlsAlaValTyThrLeuThrSerLeu 101
    |||||||
OY 359 TACCGACAAATATACAAAGTTTACTTGGAAATGAATTCAAGAGAGGAAGATGAAGTGG 418
    |||||||
Db 102 TyraAGGInTyTrhSerLeuLeuGIlysmetAsnSerGIuGIuAlaSpIIValTrp 121
    |||||||
OY 419 CAGGTGATCATATGAGCCAGACAGCTAGATGACTTCAAAAACACCAAGAGACTTGAAGCTG 478
    |||||||
Db 122 GIuValIIleIIleGlyAlaIArgAlaGIuMetThrSerTyShISGInGIuTrnLeuysLeu 141
    |||||||
OY 479 GAACCACTTGGATGACTGCACTGTGCTTTCAGAGATGGCAGCAGAACTGCATATCAA 538
    |||||||
Db 142 GIuTrnTrnTrpMetThrAlaValAlGIlyLeuSerGIuMetAlaIIaGIuAlaIArgIn 161
    |||||||
OY 539 ACTGGCGCAGATCAAGCCTCTATTAACCGCAGGAATCAATTCACATTCACCTGGTGAAC 598
    |||||||
Db 162 ThrGIyAlaAspGIuAlaSerIIleThrAlaArgAsnHISIIleGIuLeuValLysLeuGIu 181
    |||||||
OY 599 GTGGAAGAGGTGCACACAGTCTCCCGGAAGCAGAAACCAAGCTGCGAAGACACAGATA 658
    |||||||
Db 182 ValGIuGIuValaHISGIuLeuSerArgIyAlaGIuTrnThrsLeuAlaGIuIIaGIuIIle 201
    |||||||
OY 659 GAAGAGCTCCGTCAAGAAAACACAGGAGGAAAGGGGAGAGCGGCTGAGTCGAGGCAGAGAG 718
    |||||||
Db 202 GIuGIuLeuArgGIuLysThrGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 221
    |||||||
OY 719 GCCTACCTGGCTGAGGAT 736
    |||||||
Db 222 AlaTyLeuArgGIuAsp 227
    |||||||

RESULT 2
; Sequence 2, Application US/09798116
; Patent No. US20020110851A1
; GENERAL INFORMATION:
; APPLICANT: Ekerdt, Paul
; APPLICANT: Vaux, David
; TITLE OF INVENTION: NO. US20020110851A1el Polypeptides, Modulatory Agents Therefor an
; FILE REFERENCE: 10338-004US
; CURRENT APPLICATION NUMBER: US/09/798.116
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: AU PQ5995/00
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2

```

```

; LENGTH: 237
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-798-116-2

```

Alignment Scores:

Pred. No.:	7.57e-82	Length:	23
Score:	1016.00	Matches:	20
Percent Similarity:	91.21%	Conservative:	14
Best Local Similarity:	85.36%	Mismatches:	19
Query Match:	41.54%	Indels:	1
DB:	10	Gaps:	2

US-09-939-293-1 (1-1358) x US-09-798-116-2 (1-237)

```

OY 20 ATGGGGGCTGTGAGAGTGTGGCTGTGGCGGACAGCTAACTTACTTCTGAGTACAGACG 79
Db 1 MetAlaAlaLeuArgSerTTPValThrArgSerValGcSerLeuPheArgTylArgGln 20
OY 80 TGTTTGTGTGTTCCGTGTGTGGCTAACTTTAAGAAAGGGGTGTTCTCGAATTGTGTAAGA 139
Db 21 ArgPhe-----ProValLeuAlaAsnSerLysLysArgGcysPheSerGlnLeuLys 38
OY 140 CCATGGCACAACAACTGTATACGATTTGGAGTGTGAGTACCCGTGTGGCTTCCATTTGCA 199
Db 39 ProTPrhLysLysFthValLeuThrGlyPheGlyMetThrLeuGcysAlaValProIleAla 58
OY 200 CAGAAATAGAGACCCATCTCCCTTAGTGTAGTGAAGCATTTGATGAGAGAGACAGTCTTTG 259
Db 59 GlnLysSerGluProHisLysSerLeuSerAsnGlnAlaLeuMetAlaGArgAlaValSerLeu 78
OY 260 GTAACAGATAGACACCTACCTTCTCTCTGACAGCCACATATAGCCTTGATTAAGACATTT 319
Db 79 ValThrAspSerThrSerThrPheLeuSerGlnThrThrTyrAlaLeuIleIleGlnAlaLe 98
OY 320 ACTGATATACTAAGGCTGTTTATACCTTTAACTTCTCTTTTACCGACAATATATCAAGTTTA 379
Db 99 ThrGluTyrThrLysAlaValAlaTyrThrLeuValSerLeuTyrArgGlnTyrThrSerLeu 118
OY 380 CTGGGAAAATGATATTCAGAGAGAGAGAAATGAAGTGTGTGGCAGGTATATATAGAGCCAGA 439
Db 119 LeuGlnLysMetAlaSerGlnGlnGlnLysArgLysValTTPGlnValIleIleGlyAlaArg 138
OY 440 GCTGAGTACACTTCAAAACACCACAGACTACTTGAAGCTGGAAACACCTTGGATGCTGCA 499
Db 139 ValGlnMetThrSerLysGlnGlnGlnIuTyrLeuLysLeuGlnIuThrThrTTPMetThrAla 158
OY 500 GTTGTGCTTTTCAGAGATGGCAGACAGAACCTGCATATCAACTGGCGGCAGATTCAGGCTCT 559
Db 159 ValGlyLeuSerGlnMetAlaIleAlaGlnIuAlaIaTyrGlnThrGlnAlaAspIleAlaSer 178
OY 560 ATAACCCGACAGATTCACATTTACGCTGGTGAACCTGCAGGTGGGAAGAGAGTGCACACGTC 619
Db 179 IleThrAlaArgAsnHisIleGlnLeuValLysSerGlnValGlnGlnValAlaArgGlnLeu 198
OY 620 TCCCGAAGACGCAAGAACGCTGGCAGAGACACAGATAGAAAGAGCTCCGTACAGAAACA 679
Db 199 SerGlnLysAlaGlnIuThrLysLeuAlaGlnIuAlaGlnIuThrLysGlnLeuHisGlnLysAla 218
OY 680 CAGAGAGAGGGGAGACCGGCTGAGTGGAGCAGAGAGGCTTACCTCCGTGAGAGAT 736
Db 219 GlnGlnValSerAspGlnGlnIuAlaAspGlnGlnGlnIuAlaTyrLeuArgGlnLysp 237

RESULT 3
US-09-798-116-4
; Sequence 4, Application US/09798116
; Patent No. US20020110851A1
; GENERAL INFORMATION:
; APPLICANT: Verhagen, Anne Marie
; APPLICANT: Ekert, Paul
; APPLICANT: Vaux, David
; TITLE OF INVENTION: NO. US20020110851A1el Polypeptides, Modulatory Agents Therefor
; FILE REFERENCE: 10338-004US

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Alignment Scores:	
Pred. No.:	7.57e-82
Score:	1016.00
Percent Similarity:	91.21%
Best Local Similarity:	85.36%
Query Match:	41.34%
DB:	10
Length:	237
Matches:	204
Conservative:	104
Mismatches:	19
Indels:	2
Gaps:	1

US-09-939-293-1 (1-1358) x US-09-798-116-4 (1-237)

QY	20	ATGGCGGCTCTGGAAGAGTGGCTGCTCCGACGCACTAACTCATCTTCAGGTACAGACAG	79
Db	1	MetAlaAlaLeuAlaGserTTPValThrArgSerValYlsSerLeuPheAlaGTYATArgGln	20
QY	80	TGTTTGTGTGTCTCTGTTTGGCTAACTTAAAGAAGCGGTGTTCTCAGATTGATGAGA	135
Db	21	ArgPhe-----ProValLeuAlaIAsnSerLysLysArgLysPheSerGlnLeuIleLys	38
QY	140	CCATGGGCACAAAACCTGTGACGATGGCTTTGGAGTAACCCGTGTGGCGGTTCCATTTGCA	199
Db	39	ProThrPstLysTrpThrValLeuThrGlyPheGlyMetThrLeuCysAlaValProIleAla	58
QY	200	CAGAAATCAGAGCCCTCAATCCCTAGTAAAGTAACATTGATGAGAGAGAGAGCTCTTGG	255
Db	59	GlnLysSerGlnProHisSerLeuSerAsnGlnAlaLeuMetAlaGlyAlaValSerLeu	78
QY	260	GTAACAGATAGCACCCCTACCTTTCTCTCTCAGACCAACATATGGCGTATGGAAGCTATT	315
Db	79	ValThrAspSerThrSerThrPheLeuSerGlnThrThrTrpAlaLeuIleGlnAlaIle	98
QY	320	ACTGAATATATCTAAGGCTGTTTATACCTTAACCTTCTCTTAAACCAATATACAGTTTA	379
Db	99	ThrIuTrpTrpThrLysAlaValAlaTyrThrLeuValSerLeuTyrArgGlnTyrThrSerLeu	118
QY	380	CTTGGGAAAATGAATTTCAGGAGAGAAAGATAACTGTGGCGAGTGATCATATGAGACCGA	433
Db	119	LeuGlyLysMetAsnSerGlnGlnGlnLysPheValTrpGlnValIleIleGlyAlaArg	133
QY	440	GCTGAGATGACTTCAAAAACACCAAGAGTACTTGAAGCTGGGAACCACTTGGATACTGCA	495
Db	139	ValGlnMetThrSerLysGlnGlnGlnTyrLeuLysLeuGlnTyrThrTrpMetThrAla	155
QY	500	GTTGGCTTTTAAAGATGGGACGACGAAGCTGCATATCAAACTGGCGGACATCAGGCTCT	555
Db	159	ValIlyLeuSerGlnMetAlaIleAlaGlnAlaIleTyrGlnThrTrpGlyAlaAspGlnAlaSer	178
QY	560	ATAACCGGCAGAGATTCACATTCACCTGGTGAACATGCGAGGGGAAGAGTCCACACACTC	613
Db	179	IleThrAlaIleGlyAsnHisIleGlnLeuValLysSerGlnValGlnGlnValAlaArgGlnLeu	196
QY	620	TCCCGGAAAGCAGAAACCAACCTGGCAGAAAGCACAGATAGAAAGAGCTCCGACAGAAACA	679
Db	199	SerGlnLysAlaGlnTyrThrLysLeuAlaGlnIleAlaGlnThrLysGlnLeuHisGlnLysAla	218
QY	680	CAGAGCAAGGGGAGACCGGCGCTGACTCGGACGACGAGAGGCTTACCTGCGTAGAGAT	736
Db	219	GlnGlnValSerAspArgGlnGlyAlaAspGlnGlnGlnAlaIleTyrLeuArgGlnLysP	237

```

: Patent No. US20020110851A1
: GENERAL INFORMATION:
: APPLICANT: Verhagen, Anne Marie
: APPLICANT: Ekerlt, Paul
: APPLICANT: Vaux, David
: TITLE OF INVENTION: NO. US20020110851A1el Polypeptides, Modulatory Agents Therefor
: FILE REFERENCE: 10338-004US
: CURRENT APPLICATION NUMBER: US/09/798,116
: CURRENT FILING DATE: 2001-03-02
: PRIOR APPLICATION NUMBER: AU P0595/00
: PRIOR FILING DATE: 2000-03-02
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 7
: LENGTH: 202
: TYPE: PRT
: ORGANISM: Homo sapiens
: OS-09-798-116-7

```

Alignment Scores:	
Pred. No.:	8, 21e-81
Score:	1004.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	41.05%
DB:	10
US-09-939-293-1 (1-1358) x US-09-798-116-7 (1-202)	
	Length: 2020
	Matches: 2020
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0

US-09-939-293-1 (1-1358) x US-09-798-116-7 (1-202)

QY	131	TTGTAAGACCATGAGCAAAACCTGTACGATTGGCTTTGGAGTAACCCGTGGCGGTT	130
Db	1	LeuLeuArgProThrPHisLysThrValThrIleGlyHebGlyValThrLeuCysAlaVal	20
QY	191	CCATATGCACAGAAATCAGAGCCCTCATCTCCCTTAGTAGTAAGCATTTGAGAGACA	250
Db	21	ProIleLeuGlnLysSerGluProHisSerLeuSerSerGluAlaLeuMetArgAla	40
QY	251	GTGCTATTGCAACAGATAGACCTGTCACCTTCTCTCCAGACCAATATGCGTCAAT	310
Db	41	ValSerLeuValThrAspSerThrSerThrPheLeuSerGlnThrThrTyralaLeuIle	60
QY	311	GAACGCTATTACTGAATATACTAAGGCTGTTATACCTTAACCTCTCTTCCGACAATAT	370
Db	61	GluValIleThrGlnLysThrLysAlaValLysThrLeuThrSerLeuTyraArgGlnTy	80
QY	371	ACAAAGTTACTTGGGAAATATCAATTCAGAGAGAGACATGAGTGGCCAGCTGATATA	430
Db	81	ThrSerLeuLeuGlnLysMetAsnSerGlnGlnLysArgValaItrGlnValIleIle	100
QY	431	GGAGCCAGAGCTGAGATGACTTCAAAACCCAAAGATCTTGAAGCTGGAAACCACTGG	490
Db	101	GlyAlaArgAlaGluMetThrSerLysHisGlnGlnLysLeuGlnThrThrTTP	120
QY	491	ATGACTCCAGTTGGCTCTTTCAGAGATGGCAGACAGAACCTCATATCAACTGGCGCAT	550
Db	121	MetThrAlaValGlyLeuSerGluMetAlaIleGlnIleAlaIleTyrglnThrGlyAlaasp	140
QY	551	CAGGCTCTATTAACCGCCAGAAATCAATTCAGCTGGTGAACCTGCAGGTGGGAAGAGTG	610
Db	141	GlnIleSerIleThrAlaArgAsnHisIleGlnLeuValLysLeuGlnValGlnGlnVal	160
QY	611	CACACAGCTCCCGGAAAGCGAAGAACCAAGCTGGCACAAGCACACAGATAGAAGAGCTCGT	670
Db	161	HisGlnLeuSerArgLysAlaGlnThrLysLeuAlaIleGlnIleGlnGlnLeuArg	180
QY	671	CAGAAAACACAGAGAAAGGGGAGGAGCGGGCTGATGTCGAGCAGAGAGGCTTACCTGCGT	730
Db	181	GlnLysThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	200
QY	731	GAGGAT 736	
Db	201	Gluasp 202	


```

: CURRENT APPLICATION NUMBER: US/09/798.116
: CURRENT FILING DATE: 2001-03-02
: PRIOR APPLICATION NUMBER: AU P05995/00
: PRIOR FILING DATE: 2000-03-02
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 9
: LENGTH: 84
: TYPE: PRT
: ORGANISM: Rattus sp.
US-09-798-116-9

Alignment Scores:
Pred. No.: 1.08e-21 Length: 84
Score: 333.00 Matches: 68
Percent Similarity: 86.05% Conservative: 6
Best Local Similarity: 79.07% Mismatches: 10
Query Match: 13.61% Indels: 2
DB: 10 Gaps: 1

US-09-939-293-1 (1-1358) x US-09-798-116-9 (1-84)
QY 20 ATGCGCGCTGAGAGAGTGGCTGTGGCGCAGCGTAACCTTATCTCAGGTACAGACAG 79
Db 1 MetaAlaAlaLeuArgSerThrPheThrValThrPheLeuPheArgTyrGlyGln 20
QY 80 TGTGTGTGTCTCTGCTGTGGCTTAAGAACGCGTGTCTCAGATTGATAGA 139
Db 21 ArgPhe-----ProValSerIleAsnSerLysLysArgCysPheSerGluLeuIleArg 38
QY 140 CCATGGCACAACAACTGTGACGATGTGGCTTTGGAGTAACCCGTGTGGCTTCATTGCA 199
Db 39 ProThrPheLysThrMetLeuThrGlyPheGlyValThrLeuCysAlaValProIleAla 58
QY 200 CAGAAATCAGAGCCTCATTCCTTCTAGTAGTGAAGCATTGATGAGAGAGCAGTGTCTTG 259
Db 59 GlnLysSerGluProGlnSerLeuSerAsnGlnAlaLeuMetArgAlaValSerLeu 78
QY 260 GTAACAGATTACACCTCT 277
Db 79 ValThrAsnSerThrSer 84

RESULT 8
US-09-939-293-2
: Sequence 2, Application US/09939293
: Patent No. US20020132786A1
: GENERAL INFORMATION:
: APPLICANT: Alnemri, Emed S.
: TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
: FILE REFERENCE: 480140.465
: CURRENT APPLICATION NUMBER: US/09/939.293
: CURRENT FILING DATE: 2001-08-24
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 40
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-939-293-2

Alignment Scores:
Pred. No.: 3.53e-09 Length: 40
Score: 190.00 Matches: 40
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.77% Indels: 0
DB: 10 Gaps: 0

US-09-939-293-1 (1-1358) x US-09-939-293-2 (1-40)
QY 185 GCGGTTCTATTGACAGAAATCAGAGCGCTCATTCCTTAGTAGTGAAGCATTGATGAG 244
|||||
```

```

Db 1 AlaValProIleAlaGlnLysSerGluProHisSerLeuSerGlnAlaLeuMetArg 20
QY 245 AGACGAGTGTCTTTGGTAACAGATAGCACCTCTACCTTTCTCTCAGACCAATATGCG 304
Db 21 ArgAlaValSerLeuValThrAspSerThrSerThrPheLeuSerGlnThrTyrAla 40

RESULT 9
US-09-939-293-8
: Sequence 8, Application US/09939293
: Patent No. US20020132786A1
: GENERAL INFORMATION:
: APPLICANT: Alnemri, Emed S.
: TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
: FILE REFERENCE: 480140.465
: CURRENT APPLICATION NUMBER: US/09/939.293
: CURRENT FILING DATE: 2001-08-24
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 39
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-939-293-8

Alignment Scores:
Pred. No.: 7.9e-09 Length: 39
Score: 186.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.60% Indels: 0
DB: 10 Gaps: 0

US-09-939-293-1 (1-1358) x US-09-939-293-8 (1-39)
QY 185 GCGGTTCTATTGACAGAAATCAGAGCGCTCATTCCTTAGTAGTGAAGCATTGATGAG 244
Db 1 AlaValProIleAlaGlnLysSerGluProHisSerLeuSerGlnAlaLeuMetArg 20
QY 245 AGACGAGTGTCTTTGGTAACAGATAGCACCTCTACCTTTCTCTCAGACCAATAT 301
Db 21 ArgAlaValSerLeuValThrAspSerThrSerThrPheLeuSerGlnThrTyr 39

RESULT 10
US-09-939-293-11
: Sequence 11, Application US/09939293
: Patent No. US20020132786A1
: GENERAL INFORMATION:
: APPLICANT: Alnemri, Emed S.
: TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
: FILE REFERENCE: 480140.465
: CURRENT APPLICATION NUMBER: US/09/939.293
: CURRENT FILING DATE: 2001-08-24
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11
: LENGTH: 35
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-939-293-11

Alignment Scores:
Pred. No.: 6.67e-07 Length: 35
Score: 164.00 Matches: 35
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.70% Indels: 0
DB: 10 Gaps: 0

US-09-939-293-1 (1-1358) x US-09-939-293-11 (1-35)
QY 185 GCGGTTCTATTGACAGAAATCAGAGCGCTCATTCCTTAGTAGTGAAGCATTGATGAG 244
|||||
```

```
Db 1 AlavAlProileAlaGlnlySerGluProHisSerLeuSerSerGluAlaLeuMetArg 20
Qy 245 AGACAGCTGCTTGTGTAACAGATAGACACCTCTACCTTCTCTCT 289
Db 21 ArgAlaValSerLeuValThrAspSerThrSerThrPheLeuSer 35
RESULT 11
US-09-798-116-24
; Sequence 24, Application US/09798116
; Patent No. US20020110851A1
; GENERAL INFORMATION:
; APPLICANT: Verhagen, Anne Marie
; APPLICANT: Ekerl, Paul
; APPLICANT: Vaux, David
; TITLE OF INVENTION: No. US20020110851A1 Polypeptides, Modulatory Agents Therefor ar
; FILE REFERENCE: 10338-004US
; CURRENT APPLICATION NUMBER: US/09/798,116
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: AU PQ5995/00
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 32
; TYPE: PRT
; ORGANISM: synthetic
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6)..(6)
; OTHER INFORMATION: M is methionine sulfoxide
; NAME/KEY: misc_feature
; LOCATION: (14)..(14)
; OTHER INFORMATION: M is methionine sulfoxide
US-09-798-116-24
Alignment Scores:
Pred. No.: 2.68e-06 Length: 32
Score: 157.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.42% Indels: 0
DB: 10 Gaps: 0
US-09-939-293-1 (1-1358) x US-09-798-116-24 (1-32)
Qy 476 CTGGAACCACTTGATGACGACGAGTGGCTTCACAGATGAGCAGCAGAGCTGCATAT 535
Db 1 LeuGluThrThrTrpMetThrAlaValGlyLeuSerGluMetAlaAlaGluAlaLys 20
Qy 536 CAACCTGGCGCAGATCAGGCGCTCTATACCGCCAGG 571
Db 21 GlnThrGlyAlaAspGlnAlaSerIleThrAlaArg 32
RESULT 12
US-09-798-116-10
; Sequence 10, Application US/09798116
; Patent No. US20020110851A1
; GENERAL INFORMATION:
; APPLICANT: Verhagen, Anne Marie
; APPLICANT: Ekerl, Paul
; APPLICANT: Vaux, David
; TITLE OF INVENTION: No. US20020110851A1 Polypeptides, Modulatory Agents Therefor ar
; FILE REFERENCE: 10338-004US
; CURRENT APPLICATION NUMBER: US/09/798,116
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: AU PQ5995/00
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 73
; TYPE: PRT
```

```
; ORGANISM: Platichthys flesus
US-09-798-116-10
Alignment Scores:
Pred. No.: 2.02e-05 Length: 73
Score: 148.50 Matches: 30
Percent Similarity: 81.97% Conservative: 20
Best Local Similarity: 49.18% Mismatches: 8
Query Match: 6.07% Indels: 3
DB: 10 Gaps: 1
US-09-939-293-1 (1-1358) x US-09-798-116-10 (1-73)
Qy 173 GTAACCTGTGTGGCTTCCTATGACAG-----AAATCAGACCTCATTCCTT 223
Db 13 MetSerIleAlaSerLeuSerValAlaArgGlyLeuPheThrGlnGluThrLeu 32
Qy 224 AGTAGTCAGCATTTGATGAGAGAGCAGTGTCTTGGTACACATFACACCTCTACTT 283
Db 33 ThrHisAspSerLeuIleArgAlaValSerValAlaThrAspSerSerThrPhe 52
Qy 284 CTCCTCAGACCAATATGCTGATGAGCTATGAGCTATGATATGATTAAGCGTGTAT 343
Db 53 LeuSerGlnThrThrLeuAlaLeuIleAspAlaLeuThrAspTyrSerLysAlaValHis 72
Qy 344 ACC 346
Db 73 Thr 73
RESULT 13
US-09-939-293-7
; Sequence 7, Application US/09939293
; Patent No. US20020132786A1
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
; FILE REFERENCE: 480140.465
; CURRENT APPLICATION NUMBER: US/09/939,293
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-939-293-7
Alignment Scores:
Pred. No.: 6.78e-05 Length: 30
Score: 141.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.76% Indels: 0
DB: 10 Gaps: 0
US-09-939-293-1 (1-1358) x US-09-939-293-7 (1-30)
Qy 185 GCGGTCTCTATTGACAGGAATCAGACGCTCATTCCTGTAGTGAAGATTGATGAGG 244
Db 1 AlavAlProileAlaGlnlySerGluProHisSerLeuSerSerGluAlaLeuMetArg 20
Qy 245 AGACAGCTGCTTGTGTAACAGATAGACAC 274
Db 21 ArgAlaValSerLeuValThrAspSerThr 30
RESULT 14
US-09-798-116-25
; Sequence 25, Application US/09798116
; Patent No. US20020110851A1
; GENERAL INFORMATION:
; APPLICANT: Verhagen, Anne Marie
; APPLICANT: Ekerl, Paul
```



```

; APPLICANT: Vaux, David
; TITLE OF INVENTION: NO. US20020110851A1el Polypeptides, Modulatory Agents therefor ar
; FILE REFERENCE: 10338-004US
; CURRENT APPLICATION NUMBER: US/09/798,116
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: AU P05995/00
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 29
; TYPE: PRT
; ORGANISM: synthetic
US-09-798-116-25

Alignment Scores:
Pred. No.: 0.000151 Length: 29
Score: 137.00 Matches: 29
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.60% Indels: 0
DB: 10 Gaps: 0

US-09-939-293-1 (1-1358) x US-09-798-116-25 (1-29)
QY 248 GCAGTGTCTTGGTAACAGATAGCAGCTTACTTCTCTCAGACACCATATGCGTTG 307
DB 1 AAlaValSerLeuValThrAspSerThrPheLeuSerGlnThrThyrAlaLeu 20
QY 308 ATTGAAGCTATTACTGAATATATCTAAG 334
DB 21 lIleGlnAlaIleThrGlnTyrThrIlys 29

RESULT 15
US-09-764-868-701
; Sequence 701, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 701
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-701

Alignment Scores:
Pred. No.: 0.0753 Length: 401
Score: 111.00 Matches: 87
Percent Similarity: 39.50% Conservative: 54
Best Local Similarity: 24.37% Mismatches: 112
Query Match: 4.54% Indels: 104
DB: 9 Gaps: 19

US-09-939-293-1 (1-1358) x US-09-764-868-701 (1-401)
QY 311 GACGCTATTACTGAA-----TATACTAAGCT-----GTTTATACCTTA 349
DB 97 GlnThrIleThrLysGlnTyrTyrArgAlaGlnGlnIlePheLeuValTyrAspIle 116
QY 350 ACTTCCTTACCGACATATACAGTTTACTGGAAATGAAATTCAGAGAGGAA--- 406
DB 117 SerSer---GluArgSerTyrGlnHisIleMetLysTyrPvalSerAspValaspGlnTyr 135
QY 407 -----GATGAAGTGTGCAGGTGATCATAGAGCCAGAGCTGATGATCAACAAACAC 460
DB 136 AlaprogIugLysValGlnIleLysIleLeuIleGlyAsnLysAlaAspIleGlnIleLysArg 155
```

```

QY 461 CAAGAGTACTTGAACGTGGAAACACACTTGGATGACTGCAGTGGTTCCTTACAGATGGCA 520
DB 156 Gln-----ValGly---ArgGlnGlnGly 162
QY 521 GCAGAGCTGCATATCAACATGCGCAGATCAGGCTGTATACCGCAGGAATCACAAT 580
DB 153 GlnGlnLeuAlaLysGlnTyrGlyMetAspPheTyrGlnThrSerAlaCysThrAsnLeu 182
QY 581 CAGCTG-----GTGAACCTGCAGGTGCAAGAG 607
DB 183 AsnIleLysGlnSerPheThrArgLeuThrGlnLeuValLeuGlnAlaHisArgLysGln 202
QY 608 GTGCACACAGCTCTCCGGAAAGCAGAAACCAAGCTGCGAGAGACAGATCAAGAGCC 667
DB 203 LeuGlnGlyLeuArgMetArgAlaSerAsnGlnLeuAlaLeuAlaGlnLeuGlnGlu--- 221
QY 668 CGTCAGAAACACACAGAGAGAGG-----GAGAGACGGGCTGAGCTCGAGCAGAGAGCC 721
DB 222 -----GlnGlnGlnLysProGlnGlnGlyProAlaAsnSerSerLysThrCys 236
QY 722 TACCTGCGTAGAGATTGAGGCGCTG---AGCACACTGCCCTGTCTCCCACTCAG---TG 775
DB 237 Trp-CysCysValLeuTyrPglAlaProHisThrThrProLeuPheProGlnGlnAlaAr 256
QY 776 GGGAAAGCAGGGGCGAGATGCCACCTGCCAGGTTGGCATCTGTCTGTC----- 828
DB 256 gGlyGlnThrGlyGlnProGlyLeuCysProAla---AlaValLeuSer-CysAspAspP 275
QY 829 --ACGAGAGAGAGGGGCGAGTCTGCCCTGCACATCAGGCGAGAGCGCTTGTGAGCT 886
DB 275 rolleGlnTyrGlnLysProLeuLeuProLeuPro----- 286
QY 887 GTGAGTGCCTCCTGTGTGCTCAGGCTTGCGCTGACCT-----GATTCCTAGCCCTGG 940
DB 287 -----GlyProGlnSerGlySerAlaValIles 296
QY 941 GCAGTGCACCCCTGT-----TTACATTTACACCCA 970
DB 296 erSerSerProCysProGlnProValHisProGlyValGlyPheSerLeuPheProG 316
QY 971 CTCTGTACAGCTGCTCTTACCCATTTTCTTACCTCACACCCAAAGCATTTTGCCTACT 1030
DB 316 lnpProGlnAlaCysTyrAspProHisAspValPProGlnAlaLeuSerHisHisProAlaP 336
QY 1031 GGGTCAGAGAGAGG---AGTCCCTTTTTCATGACCCCTTAAGTTCAGCAACTGTTAACT 1087
DB 336 roThrArgGlnGlnProGly-LeuGlnSerArgProLeuSerAlaAlaPro-PheSerVa 355
QY 1088 GTTT-----TCAGTCTTATTACGTCGTCGCAAAATGATTTAGTACTTGTCCCTCT 1138
DB 355 lHisArgValSerSerLeuLeuPheLeuSerSerProThrSerLeuSerLeuThrProPr 375
QY 1139 GTTGGAGATGCCAGTGTGCGACAGGGGAGGAGGCAACCTGTCACATTTGTACGATTTCTTGT 1198
DB 375 oleuArgCysValSerTyrGlnSerSerAsn----- 386
QY 1199 ATGATTTTCTGATGTGTTCTGATCGTGGCCCACTGTCCTGT 1241
DB 387 -----ProValProArgValSerCys 393
```

Search completed: February 20, 2003, 15:15:50
Job time : 19.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 20, 2003, 14:47:23 : Search time 21.5 seconds

(without alignments)
5239.521 Million cell updates/sec

Title: US-09-939-293-1

Perfect score: 2446

Sequence: 1 ggcgtccgcgcgcgcacaa.....ggacttaacacagaanaaaa 1338

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+np2_model -DEV=xip
-O=/cgn2.1/USPPO_pool/US09939293/runat_20022003.111510.14829/app-query.fasta.1.1543
-DB=SwissProt.40 -OFMT=fasta -SUFFIX=rsp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIG=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09939293_cgn2.1.130.0@runat_20022003.111510.14829 -NCPD=6 -ICPD=3
-NO_XIPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1201	49.1	239	1	SMAC_HUMAN
2	1016	41.5	237	1	SMAC_MOUSE
3	125.5	5.2	502	1	WASP_HUMAN
4	112	4.6	360	1	DMD_CHICK
5	111	4.6	1323	1	NME4_MOUSE
6	111	4.6	1323	1	NME4_RAT
7	110	4.5	589	1	SPY_DROME
8	107.5	4.4	1520	1	TOP2_CAEEL
9	106.5	4.4	3164	1	TROU_HSV1
10	106	4.4	520	1	WASP_MOUSE
11	104.5	4.3	1219	1	JAG1_RAT
12	103.5	4.2	4684	1	PLEI_HUMAN
13	103	4.2	592	1	LAM2_MOUSE
14	103	4.2	1433	1	REST_CHICK
15	102.5	4.2	3924	1	ANK2_HUMAN
16	102.5	4.2	5376	1	ZAN_MOUSE
17	102	4.2	2116	1	MYO2_DICDI
18	101.5	4.2	500	1	FXE2_HUMAN

19	101	4.1	502	1	K2M3_SHEEP	P25691 ovls aries
20	101	4.2	1509	1	GSRL_HUMAN	Q9uzm4 homo sapien
21	100	4.1	4377	1	ANK3_HUMAN	Q12955 homo sapien
22	99.5	4.1	1087	1	NFR_MOUSE	P19246 mus musculu
23	99.5	4.1	1213	1	FMN_CHICK	Q05858 gallus galli
24	99	4.1	616	1	REPX_HUMAN	P48382 homo sapien
25	99	4.1	721	1	YR82_MYCTU	Q10690 mycobacteri
26	99	4.1	1336	1	NME4_HUMAN	Q15399 homo sapien
27	98.5	4.1	988	1	CIC1_HUMAN	P35523 homo sapien
28	98	4.0	548	1	NFL_PIG	P02347 sus scrofa
29	98	4.0	4427	1	PKSL_BACSU	Q05470 bacillus su
30	97.5	4.0	975	1	CUT1_CANFA	P39881 canis fam1
31	97.5	4.0	4473	1	PLE1_CRIGR	Q91155 cricetus
32	97	4.0	603	1	NMEL_MOUSE	O08919 mus musculu
33	97	4.0	831	1	NFR_RAT	P16884 rattus norv
34	97	4.0	4687	1	PLE1_RAT	P30427 rattus norv
35	96.5	4.0	189	1	AMEL_PIG	P45561 sus scrofa
36	96.5	3.9	331	1	MAN1_MOUSE	Q9wu40 mus musculu
37	96.5	3.9	651	1	VU47_HSV6U	O06093 human hepe
38	96.5	3.9	1043	1	FPR1_DROME	P33244 drosophila
39	96	4.0	633	1	LAI7_YEAST	Q12446 saccharomyc
40	96	4.0	1183	1	DRPL_RAT	P54258 rattus norv
41	95.5	3.9	460	1	MYCN_MARMO	Q61976 marmota mon
42	95	3.9	544	1	NFL_XENLA	P35616 xenopus lae
43	95	3.9	817	1	VRP1_YEAST	P37370 saccharomyc
44	95	3.9	1208	1	RCO4_HUMAN	O94761 homo sapien
45	94.5	3.9	503	1	WALP_HUMAN	O43516 homo sapien

ALIGNMENTS

RESULT 1

SMAC_HUMAN	STANDARD:	PRT:	239 AA.
ID	SMAC_HUMAN	Q9NR28; Q9BT11; Q9HAV6; Q96LW0;	
AC	Q9NR28; Q9BT11; Q9HAV6; Q96LW0;		
DT	16-OCT-2001 (Rel. 40; Created)		
DT	16-OCT-2001 (Rel. 40; Last sequence update)		
DT	15-JUN-2002 (Rel. 41; Last annotation update)		
DE	Smac protein, mitochondrial precursor (Second mitochondria-derived activator of caspase) (Direct IAP binding protein with low pI).		
GN	SMAC OR DIABLO.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM 1), PARTIAL SEQUENCE, FUNCTION, AND TISSUE SPECIFICITY.		
RP	MEDLINE=20383536; PubMed=10929711;		
RX	Du C., Fang M., Li Y., Li L., Wang X.;		
RA	"Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition."		
RT	Cell 102:33-42(2000).		
RL	[2]		
RN	SEQUENCE FROM N.A. (ISOFORM 1).		
RP	Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,		
RA	Suzuki Y., Ohtsuyashi M., Nishi T., Shibaara T., Tanaka T.,		
RA	Nakamura Y., Isogai T., Sugano S.;		
RT	"NEO human cDNA sequencing project."		
RL	Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.		
RP	PubMed=10950947;		
RA	Sridivasula S.M., Delta P., Fan X.J., Fernandes-Alnemri T., Huang Z.,		
RA	Alnemri E.S.;		
RT	"Molecular determinants of the caspase-promoting activity of Smac/DIABLO and its role in the death receptor pathway."		
RL	J. Biol. Chem. 275:36152-36157(2000).		
RN	[4]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RP	TISSUE=Cerebellum;		
RC	Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Suzuki Y.,		
RA	Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,		

RA	Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,
RA	Karakami B., Nagai K., Isgai T., Sano S.;
RT	"NEDO human cDNA sequencing project.";
RL	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
RN	[5]
RP	SEQUENCE FROM N.A. (ISOFORM 1).
RC	TISSUE=Muscle, and Uterus;
RA	Strausberg R.;
RL	Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
RN	[6]
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 56-239.
RX	MEDLINE=20426096; PubMed=10972280;
RA	Chai J., Du C., Wu J.W., Klyn S., Wang X., Shi Y.;
RT	"Structural and biochemical basis of apoptotic activation by
RT	Smac/DIABLO.";
RL	Nature 406:855-862(2000).
RN	[7]
RP	STRUCTURE BY NMR OF 56-64 IN COMPLEX WITH BIRC4.
RX	MEDLINE=21202061; PubMed=11140637;
RA	Liu Z., Sun C., Olejniczak E.T., Meadows R.P., Betz S.F., Oost T.,
RA	Herrmann J., Wu J.C., Fesik S.W.;
RT	"Structural basis for binding of Smac/DIABLO to the XIAP BIR3
RT	domain.";
RL	Nature 408:1004-1008(2000).
CC	-I- FUNCTION: PROMOTES APOPTOSIS BY ACTIVATING CASPASES IN THE
CC	CYTOCHROME C/APAF-1/CASPASE-9 PATHWAY. ACTS BY OPPOSING THE
CC	INHIBITORY ACTIVITY OF INHIBITOR OF APOPTOSIS PROTEINS (IAP).
CC	-I- SUBUNIT: Homodimer. Interacts with BIRC2, BIRC3, BIRC4/XIAP and
CC	BIRC7.
CC	-I- SUBCELLULAR LOCATION: MITOCHONDRIAL BUT RELEASED INTO THE CYTOSOL
CC	WHEN CELLS UNDERGO APOPTOSIS.
CC	-I- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/Diablo-S;
CC	are produced by alternative splicing.
CC	-I- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED WITH HIGHEST EXPRESSION
CC	IN TESTIS. EXPRESSION IS ALSO HIGH IN HEART, LIVER, KIDNEY,
CC	SPLEEN, PROSTATE AND OVARY. LOW IN BRAIN, LUNG, THYMUS AND
CC	PERIPHERAL BLOOD LEUKOCYTES.
CC	-I- DOMAIN: The mature N-terminus mediates interaction with
CC	BIRC4/XIAP.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AF262240; AAF87716.1; -
DR	EMBL; AK024768; BAB14994.1; -
DR	EMBL; AF298770; AAG22077.1; -
DR	EMBL; AK057778; BAB171568.1; -
DR	EMBL; BC004417; AAH04417.1; -
DR	PDB; 1FEW; 13-SEP-00.
DR	PDB; 1G3F; 10-JAN-01.
DR	MIM; 605219; -
KW	Transit peptide; Mitochondrion; Apoptosis; Alternative splicing;
KW	3D-structure.
FT	TRANSIT 1 55 MITOCHONDRION.
FT	CHAIN 56 239 SMAC PROTEIN.
FT	SITE 56 60 IAP-BINDING MOTIF (BY SIMILARITY).
FT	VARSPLIC 1 60 MAALSWLSRVSSTFFPRROGLCPVAVANFKKFSLEIRP
FT	WHKRWTFQVTLCAVPIA -> MKSPYF (IN
FT	ISOFORM 2).
FT	K -> E (IN REF. 4).
FT	K -> R (IN REF. 2).
FT	MISSING (IN REF. 4).
FT	E -> K (IN REF. 4).
FT	SEQUENCE 239 AA; 27131 MW; 70C2AE0DC654D031 CRC64;

Alignment Scores:	
Pred. No.:	1.51e-80
Score:	1201.00
	Length: 239
	Matches: 239

```
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 49.10% Indels: 0
DB: 1 Gaps: 0
```

US-09-939-293-1 (1-1358) x SMAC_HUMAN (1-239)

OY	20	ATGGCGGCTCTGAAAGAGTTGGCGTCGGCCACCGACCAATCTATCTTCAGATCAGACG	79
Db	1	Met1a1a1a1eulysser1trpleuser1argserv1ltnrser1phe1a1g1y1r1a1g1n	20
OY	80	TGTTTGTGTCTTCCTGTTGTGGCTAACTTTAAGAAGCGGTCTTTCTCAGATTGATAAGA	13
Db	21	Cys1eucy1val1Proval1a1a1asn1p1e1y1s1a1r1c1y1s1Phe1ser1G1u1eul1e1a1r	40
OY	140	CCATGGACCAAAAC1CTGACGATGGCTTTGGAGTAAACCCGTGGGGTTCCTATTGGCA	19
Db	41	Pro1p1a1s1y1s1thr1Val1ltn1r1e1g1y1Phe1g1y1a1ltn1r1eucy1s1a1a1a1Pro1l1e1a1	60
OY	200	CAGAAATCAGAGCCTCATTCCTTACCTTAAGTATGAGCATTTGATGAGAGAGCATGTCTTTTG	25
Db	61	G1n1ys1ser1G1u1Pro1h1s1ser1Leu1ser1ser1G1a1a1e1u1Met1a1r1a1a1Val1ser1e1u	80
OY	260	GTAACCAATAGCACCTCTACCTTTCTCTCTCAGACCACATATGGCTTGATTGAACCTATT	31
Db	81	Val1ltn1r1asp1er1thr1ser1ltn1r1Phe1Leu1ser1G1n1ltn1r1y1a1a1eul1e1g1a1a1l1e1	10
OY	320	ACTGAATATCTTAAGCTGTTTATACCTTAACCTCTCTTTCACCGACAATATACAGTTTA	37
Db	101	Thr1u1y1r1thr1r1ys1a1a1a1y1r1h1r1euth1r1ser1eul1y1a1r1g1n1y1r1h1ser1e1u	12
OY	380	CTTGCGCAAAATGAATTCAGAGAGGAGATGAAGTGTGGCGAGTGATCATAGACCCAGA	43
Db	121	Leu1e1y1s1Met1asn1ser1G1u1G1n1u1s1p1a1r1G1n1a1l1r1e1e1g1a1a1a1r1y	14
OY	440	GCTGAGATGACTTCAAAACACCAAGAGTCTTGAAGCTGGAAACCACTTGATGAGACTGCA	49
Db	141	Ala1e1u1Met1hr1ser1y1s1h1s1e1n1u1y1r1eul1y1s1eul1G1u1h1r1h1r1p1Met1h1r1a1	16
OY	500	GTGTGCTTTTCAGAGATGGCAGACAGAGCTGCATATCAATCAACGTGGCGCATCAGGCTCT	55
Db	161	Val1y1Leu1ser1ltn1Met1a1a1a1a1a1a1a1a1a1y1r1G1n1h1r1g1y1a1a1a1s1p1n1a1a1ser	18
OY	560	ATAACGGCCAGGAATACATTCAGCTGGGGAACGCGAGGGAAGAGTGGCCACGATC	61
Db	181	l1e1h1r1a1a1r1a1r1asn1h1s1t1e1G1n1eul1Val1y1s1eul1G1n1a1l1G1u1a1l1h1s1G1n1e1u	20
OY	620	TCCCGAACAACACAACCAAGCTGGCAGAGACACAGATAGAGAGAGCCCTCAGAAACA	67
Db	201	Ser1a1r1y1s1a1a1G1u1h1r1y1s1eul1a1a1a1G1n1a1a1G1n1l1e1G1u1eul1a1r1G1n1y1s1thr	22
OY	680	CAGAGAGAGGAGGAGAGCGGGCTGATGTGGACGACGAGAGGCTTCCTGCTGAGAGAT	73
Db	221	G1n1G1u1G1n1eul1y1G1u1a1r1a1G1a1u1s1er1G1u1G1n1a1a1y1r1eul1a1r1G1u1a1s1p	23
RESULT 2			
SMAC_MOUSE			
ID	SMAC_MOUSE	STANDARD:	PRT: 237 AA.
AC	Q9UIQ3; Q9CZD1; Q9DCD3;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Smac protein, mitochondrial precursor (Second mitochondria-derived		
DE	activator of caspase) (Direct IAP binding protein with low pI).		
GN	SMAC OR DIABLO.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RM	[1]		
RP	SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE		
RP	SPECIFICITY.		
RC	STRAIN=BAIB/C; TISSUE=Kidney;		

RP SEQUENCE FROM N.A.
 RX MEDLINE=95273432; PubMed=7753869;
 RA Kwan S.-P., Hagemann T.L., Radtke B.E., Blaese R.M., Rosen F.S.;
 RT "Identification of mutations in the Wiskott-Aldrich syndrome gene and
 RT characterization of a polymorphic dinucleotide repeat at DXS6940,
 RT adjacent to the disease gene.";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:4706-4710(1995).
 RL [14]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99167346; PubMed=10066431;
 RA Hagemann T.L., Kwan S.-P.;
 RT "The identification and characterization of two promoters and the
 RT complete genomic sequence for the Wiskott-Aldrich syndrome gene.";
 RL Biochem. Biophys. Res. Commun. 256:104-109(1999).
 RP [15]
 RA Blechschmidt K., Nyakatura G., Strom T.M., Drescher B., Menzel U.,
 RA Meindl A., Rosenthal A.;
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 RN [16]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN [17]
 RP VARIANTS WAS K-31; M-75; P-82; C-86; H-86; C-97; K-133; E-476 AND
 RP H-30 DEL.
 RX MEDLINE=96133285; PubMed=8528198;
 RA Kolluri R., Shehabeldin A., Peacocke M., Lamhonwah A.-M.,
 RA Teichert-Kuliszewska K., Weissman S.M., Simunovich K.A.;
 RT "Identification of WASP mutations in patients with Wiskott-Aldrich
 RT syndrome and isolated thrombocytopenia reveals allelic heterogeneity
 RT at the WAS locus.";
 RL Hum. Mol. Genet. 4:1119-1126(1995).
 RN [18]
 RP VARIANTS WAS F-27; I-48; M-75; L-86; H-86; K-131; C-187 AND K-477.
 RX MEDLINE=96133286; PubMed=8528199;
 RA Derry J.M.J., Kerns J.A., Weinberg K.I., Ochs H.D., Volpini V.,
 RA Estivill J., Walker A.P., Francke U.;
 RT "WASP gene mutations in Wiskott-Aldrich syndrome and X-linked
 RT thrombocytopenia.";
 RL Hum. Mol. Genet. 4:1127-1135(1995).
 RN [19]
 RP VARIANTS WAS VAL-56 AND GLU-236.
 RX MEDLINE=95315993; PubMed=7795848;
 RA Villa A., Notarangelo L., Macchi P., Mantuano E., Cavaign G.,
 RA Brugnoni D., Strina D., Patrosso M.C., Ramenghi U., Sacco M.G.,
 RA Ugazio A., Vezzoni P.;
 RT "X-linked thrombocytopenia and Wiskott-Aldrich syndrome are allelic
 RT diseases with mutations in the WASP gene.";
 RL Nat. Genet. 9:414-417(1995).
 RN [10]
 RP VARIANTS WAS TRP-43; MET-45; MET-75 AND CYS-86.
 RX MEDLINE=97272113; PubMed=9126958;
 RA Remold-O'Donnell E., Cooley J., Shcherbina A., Hagemann T.L.,
 RA Kwan S.-P., Kenney D.M., Rosen F.S.;
 RT "Variable expression of WASP in B cell lines of Wiskott-Aldrich
 RT syndrome patients.";
 RL J. Immunol. 158:4021-4025(1997).
 RN [11]
 RP VARIANTS WAS LYS-31 AND MET-45.
 RX MEDLINE=97253413; PubMed=9098856;
 RA Ariga T., Yamada M., Sakiyama Y.;
 RT "Mutation analysis of five Japanese families with Wiskott-Aldrich
 RT syndrome and determination of the family members' carrier status
 RT using three different methods.";
 RL Pediatr. Res. 41:535-540(1997).
 RN [12]
 RP VARIANTS WAS MET-75; LEU-84; ASP-89 AND LYS-133.
 RX MEDLINE=98350091; PubMed=9683546;
 RA MacCarthy-Morrogh L., Gaspar H.B., Wang Y.-C., Katz F., Thompson L.,
 RA Layton M., Jones A.M., Kinnon C.;
 RT "Absence of expression of the Wiskott-Aldrich syndrome protein in

RT peripheral blood cells of Wiskott-Aldrich syndrome patients.";
 RL Clin. Immunol. Immunopathol. 88:22-27(1998).
 RN [13]
 RP VARIANTS WAS VAL-56.
 RX MEDLINE=96376986; PubMed=9713366;
 RA Facchetti F., Blanzuoli L., Verml W., Notarangelo L.D., Gilliani S.,
 RA Fiorini M., Fasth A., Stewart D.M., Nelson D.L.;
 RT "Defective actin polymerization in EBV-transformed B-cell lines from
 RT patients with the Wiskott-Aldrich syndrome.";
 RL J. Pathol. 185:99-107(1998).
 RN [14]
 RP VARIANTS WAS LYS-133.
 RX MEDLINE=98092074; PubMed=9445409;
 RA Parolini O., Reesmann G., Haas O.A., Pawlowsky J., Gadner H.,
 RA Knapp W., Holter W.;
 RT "X-linked Wiskott-Aldrich syndrome in a girl.";
 RL New Engl. J. Med. 338:291-295(1998).
 RN [15]
 RP VARIANTS WAS ARG-73; MET-75; CYS-83; CYS-86 AND LYS-133.
 RX MEDLINE=99374925; PubMed=10447259;
 RA Lemahieu V., Gastier J.M., Francke U.;
 RT "Novel mutations in the Wiskott-Aldrich syndrome protein gene and
 RT their effects on transcriptional, translational, and clinical
 RT phenotypes.";
 RL Hum. Mutat. 14:54-66(1999).
 RN [16]
 RP VARIANTS WAS HIS-52 AND TRP-70.
 RX MEDLINE=2652748; PubMed=11793485;
 RA El-Hakeh J., Rosenzweig S., Oleastro M., Basack N., Berozdnlk L.,
 RA Molina F., Rivas E.M., Zelazko M., Danielian S.;
 RT "Wiskott-Aldrich syndrome in Argentina: 17 unique, including nine
 RT novel, mutations.";
 RL Hum. Mutat. 19:186-187(2002).
 CC -1- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
 CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION.
 CC -1- SUBUNIT: BINDS TO CD34, RAC, NCK, FYN, SRC KINASE FGR, BTK, ABL,
 CC WIP, AND TO THE P85 SUBUNIT OF PLC-GAMMA.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE THYMUS. ALSO
 CC FOUND TO A MUCH LESSER EXTENT, IN THE SPLEEN.
 CC -1- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH
 CC LIGAND.
 CC -1- DISEASE: DEFECTS IN WAS ARE THE CAUSE OF WISKOTT-ALDRICH SYNDROME
 CC (WAS), AN X-LINKED RECESSIVE IMMUNODEFICIENCY CHARACTERIZED BY
 CC ECZEMA, THROMBOCYTOPENIA, RECURRENT INFECTIONS, AND BLOODY
 CC DIARRHEA. DEATH USUALLY OCCURS BEFORE AGE 10.
 CC -1- DISEASE DEFECTS IN WAS ARE THE CAUSE OF ISOLATED X-LINKED
 CC THROMBOCYTOPENIA (XLT). XLT IS CLINICALLY MILD WITH SMALL
 CC PLATELETS AND SUBCLINICAL LEUKOCYTE ABNORMALITIES.
 CC -1- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U12707; AAA62663.1; ALT_SEQ.
 DR EMBL: U18935; AAA60381.1; -.
 DR EMBL: U19927; AAC50140.1; -.
 DR EMBL: AF115549; AAD26691.1; -.
 DR EMBL: AF196970; AAR06804.1; -.
 DR EMBL: BC002961; AAH02961.1; ALT_INIT.
 DR EMBL: BC012738; AAH12738.1; -.
 DR Genew: HGNC:12731; WAS.
 DR MIM: 300392; -.
 DR MIM: 301000; -.
 DR InterPro: IPR000095; PAKbox/Rhobinding.
 DR InterPro: IPR000697; RanBP1_WASP.
 DR InterPro: IPR001960; WH1.
 DR InterPro: IPR003124; WH2.


```

DR InterPro; IPR001589; Actbind_actuin.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR002349; MW.
DR InterPro; IPR001202; MW_Rsp5_WMP.
DR InterPro; IPR000433; Znf_ZZ.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00397; MW; 1.
DR Pfam; PF00435; spectrin; 22.
DR Pfam; PF00569; Zz; 1.
DR PRINTS; PR00403; WMDOMAIN.
DR SMART; SM00035; CH; 2.
DR SMART; SM00150; SPEC; 21.
DR SMART; SM00456; MW; 1.
DR SMART; SM00291; Znf_ZZ; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS50020; ACTININ_2; 1.
DR PROSITE; PS01159; MW_DOMAIN_1; 1.
DR PROSITE; PS50020; MW_DOMAIN_2; 1.
DR PROSITE; PS01357; ZF_ZZ_1; 1.
DR PROSITE; PS0135; ZF_ZZ_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; Zinc-finger.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 19 123 CH 1.
FT DOMAIN 138 241 CH 2.
FT REPEAT 341 449 SPECTRIN 1.
FT REPEAT 450 558 SPECTRIN 2.
FT REPEAT 561 669 SPECTRIN 3.
FT REPEAT 721 830 SPECTRIN 4.
FT REPEAT 832 936 SPECTRIN 5.
FT REPEAT 945 1047 SPECTRIN 6.
FT REPEAT 1050 1156 SPECTRIN 7.
FT REPEAT 1159 1265 SPECTRIN 8.
FT REPEAT 1268 1369 SPECTRIN 9.
FT REPEAT 1470 1570 SPECTRIN 10.
FT REPEAT 1573 1678 SPECTRIN 11.
FT REPEAT 1681 1782 SPECTRIN 12.
FT REPEAT 1879 1981 SPECTRIN 13.
FT REPEAT 2013 2103 SPECTRIN 14.
FT REPEAT 2106 2211 SPECTRIN 15.
FT REPEAT 2214 2321 SPECTRIN 16.
FT REPEAT 2472 2574 SPECTRIN 17.
FT REPEAT 2577 2683 SPECTRIN 18.
FT REPEAT 2686 2799 SPECTRIN 19.
FT REPEAT 2802 2904 SPECTRIN 20.
FT REPEAT 2906 2928 SPECTRIN 21.
FT REPEAT 2931 3037 SPECTRIN 22.
FT REPEAT 3052 3085 MW.
FT DOMAIN 3304 3351 ZF-TYPE.
FT ZN_FING 1171 1171 MISSING.
FT VARIANT 1669 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;

Alignment Scores:
Pred. No.: 1.22 Length: 3660
Score: 112.00 Matches: 56
Percent Similarity: 40.54% Conservative: 49
Best Local Similarity: 21.62% Mismatches: 112
Query Match: 4.58% Indels: 42
DB: 1 Gaps: 8

US-09-939-293-1 (1-1358) x KMD_CHICK (1-3660)
QY 5 TCCTGGCGGCTGCACATATGGCGGCTGTGAGAGCTGTGCGGAGCGTAACCTTCATTC 64
DB 1163 SerLeuArgLysAspLeuSerGluMetHisGluTrpIleThrGlnAlaGluGluGluTrp 1182
QY 65 TTCAGGTACAGACAGACTGTTTGTTGTTCTCTGTGGCGTAACCTTTAAAGAGAGCGGTGTTTC 124
DB 1183 LeuGlu-----ArgAspPheGluTrpLysThrPro 1192

```

```

QY 125 TCAGATTGATTAAGACCATGCGACAAAACCTGTGAGATTGGCTTTGAGTAACCTGTGT 184
DB 1193 GluGluLeuGlnLys----- 1197
QY 185 GCGGTTCCATTATTCACAGAAATCAGAGCCCTCATTCCTTAGTAGTGAACATTGATGAG 244
DB 1198 AlaValGluGluLeuLysArgLysGlnAspAlaMetGlnLysGluVal----- 1214
QY 245 AGAGCAGTCTTTTGGTAAACAGATAGCACCTTACTTCTTCCTCCACAGCCATATGCG 304
DB 1215 ---LysValLysLeuIleThrAspSerValAsnPheIleAlaLysAlaProProAla 1233
QY 305 TTGATTGAAGCTATTACTGATATATCTAAGGCTGTTTATACCTTACTTCTCTTACCGA 364
DB 1234 AlaAsnGluAlaLeuLysLysGluLeuAspValLeu-----IleThrSerTyrGlnArg 1251
QY 365 CAATATACAAAGTTTACTTGGGAAATGAAATTCAGAGAGGAGAGATGAAGTGTGGCAGGTG 424
DB 1252 LeuCysSerArgLeuAsnGlnLysCysLysThrLeuGlu-----GluValTrp----- 1267
QY 425 ATCATAGAGAGCCAGAGCTGAGATGACTTCAAAACACCAAGAGTACTTGAAGCTGGAAC 484
DB 1268 -----AlaCysTrpHisGluLeuLeuSer-----TyrLeuAspAlaGluAsn 1281
QY 485 ACTTGATGATGCTGAGATTGGTGTCTTCAGAGATGCGCAGAAAGCTGATATCAACTGGC 544
DB 1282 LysTrpLeuAsnGluValGluLeuLysLeuLysAlaIleThrGlu---AsnIleGlnLysGly 1300
QY 545 GCAGATCAGGCGCTCTATACCGCCAGATCATCATTCAGCTGAGTGAACCTGCAGGTGGA 604
DB 1301 AlaGluGluIleSerGlnSerLeuAspSerLeuGluArgLeuMetArgHisProGluAsp 1320
QY 605 GAGGTGCACCACTCTCCCGGAAGCAAGCAACCAAGCTGGCGGAGACAGATTAAGAG 664
DB 1321 AsnArgAsnGlnIleArgGluLeuAlaGlnTrpIleThrAspGlyGlyTleuLysPssu 1340
QY 665 CTCCTCCGTAACAAACACAGAGAGGAGGAGGAGCGGCTGATCGAGCAGAGAGCCTTAC 724
DB 1341 LeuIleAsnGlnLysLeuGlnLysPheAsnThrArgTrpGluGlnLeuGlnGlnIleAla 1360
QY 725 CTGCGTGAAGATTGAGGCGCTGAGCAGACACTGCCCTGTCTCCCGCATCAGTGGGAAA 781
DB 1361 ValArgArgGlnLysSerLeuGlnLysIleGlnSerAlaGlnLysThrAspLys 1379

RESULT 5
ID NME4_MOUSE STANDARD; PRT; 1323 AA.
AC Q03391;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
DE D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D).
GN GRIN2D.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93050214; PubMed=1385220;
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RA Inoue Y., Mishina M.,
RT "Cloning and expression of the epsilon 4 subunit of the NMDA receptor
RT channel.";
RL FEBS Lett. 313:34-38(1992).
RP REVISIONS.
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RA Inoue Y., Mishina M.;
RA Submitted (May-1998) to the EMBL/GenBank/DBJ databases.

```


Db	1044	Ala	heliu	sp	glu	ser	pro	pro	ala	pro	ser	arg	trp	pro	arg	ser	as	pro	glu	ser	106				
QY	765	-----	-GGACAGACAGG	CGA	GTC	GTC	CTAG	AGCC	CCCTCAAT	CTCTCA	CGACGAG	GTAA	GGCC	718											
Db	1064	Gln	pro	leu	thr	leu	thr	gly	gly	ala	-----	Gly	gly	pro	ser	arg	ala	gly	-Ala	pro	thr	ala	1081		
QY	717	TTC	GCT	CGC	GAC	AC	CGCC	CCCT	CTCT	CCCC	-----	CTT	CCCT	CC	G	T	G	T	T	676					
Db	1081	ap	ro	pro	arg	arg	arg	thr	ala	pro	pro	cys	ala	tyr	leu	as	par	leu	glu	ser	pro	ser	1101		
QY	675	TTT	CGAC	-----	-----	-----	-----	-----	-----	-----	-----	GGAC	GTC	TTT	ATC	TG	CTT	CTG	CCAGC	640					
Db	1101	ose	as	par	ser	glu	as	par	ser	glu	ser	leu	glu	tyl	ala	as	leu	glu	tyl	ala	as	leu	glu	1122	
QY	639	TTG	GTT	CTG	-----	-----	-----	-----	-----	-----	-----	CTT	CCG	GAG	AGC	GTC	GTG	CACT	-----	606					
Db	1121	p	trp	phe	ala	as	par	he	pro	tyr	pro	tyr	ala	glu	arg	leu	glu	tyl	pro	pro	glu	arg	tyl	1141	
QY	605	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	CTT	CCAC	CTG	CC	595									
Db	1141	t	trp	ser	val	as	par	lys	leu	glu	tyl	trp	arg	ala	glu	tyl	ser	trp	as	par	tyl	as	leu	pro	1161
QY	594	AGT	TTC	AC	CA	CA	CA	CT	-----	-----	-----	CAAT	GTC	AT	CTC	GCG	GTC	TTA	GAG	CC	TG	CA	CT	547	
Db	1161	g	ly	glu	tyl	pro	ala	thr	thr	thr	his	cys	arg	his	cys	ala	as	leu	glu	tyl	ala	as	leu	glu	1181
QY	546	GC	CGC	AG	CTT	GAT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	1191	
Db	1181	gh	is	-----	-----	-----	-----	-----	-----	-----	-----	leu	ser	cys	ser	his	as	glu	tyl	ala	as	leu	glu	tyl	1191
QY	486	G	TG	G	T	T	C	C	A	477	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1191
Db	1194	p	trp	ala	pro	1197																			1197

RESULT 6

NME4_RAT

ID NME4_RAT STANDARD; PRT; 1323 AA.

AC Q62645; Q63381; Q63382; Q63729; Q63730;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl D-aspartate receptor subtype 2b) (NR2D) (MMDAR2D).

GN DNR2D.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 1265-1356 FROM N.A. (ISOFORM 2).

RC STRAIN=Sprague-Dawley; TISSUE=Forebrain;

RX MEDLINE=93155102; PubMed=8428958;

RA Ishii T., Moriyoshi K., Sugihara H., Sakurada K., Kadotani H., Yoko M., Nakazawa C., Shigemoto R., Mizuno N., Masu M., Nakanishi S.; "Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits.";

RT J. Biol. Chem. 268:2836-2843(1993).

RL [2]

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Brain;

RX MEDLINE=94206533; PubMed=7512349;

RA Monyer H., Burnashev N., Laurie D.J., Sakmann B., Seeburg P.H.; "Developmental and regional expression in the rat brain and functional properties of four NMDA receptors.";

RT Neuron 12:529-540(1994).

RL [3]

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC STRAIN=Sprague-Dawley; TISSUE=Brain;

RA Boulter J., Pecht G.;

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS WITH HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT

RT "Sprouty encodes a novel antagonist of FGF signaling that patterns
RT apical branching of the Drosophila airways.",
RL Cell 92:253-263(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazek R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-Y., Andrews-Plankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Botshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garz N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jostali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinhart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpston M., Skupski M.P., Smith T.,
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Welstock G.M., Weissbach J.,
RA Williams S.M., Woodgate J., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zheng H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
[3]
RN FUNCTION;
RP MEDLINE=99387981; PubMed=10457022;
RA Reich A., Sapir A., Shilo B.-Z.;
RT "Sprouty is a general inhibitor of receptor tyrosine kinase
RT signaling";
RL Development 126:4139-4147(1999).
[4]
RN FUNCTION;
RX MEDLINE=99244704; PubMed=10226010;
RA Kramer S., Okabe M., Hachon N., Krasnow M.A., Hiromi Y.;
RT "Sprouty: a common antagonist of FGF and EGF signaling pathways in
RT *Drosophila*";
RL Development 126:2515-2525(1999).
[5]
RN FUNCTION, AND SUBCELLULAR LOCATION;
RX MEDLINE=99189751; PubMed=10089881;
RA Casci T., Vinos J., Freeman M.;
RT "Sprouty, an intracellular inhibitor of Ras signaling.";
RL Cell 96:655-665(1999).
-1- FUNCTION: INHIBITOR OF TRACHEAL BRANCHING THAT RESTRICTS BRANCH
CC BODDING BY ANTAGONIZING THE BNL-FGF PATHWAY (BNL: BRANCHLESS, AN
CC FGF INDUCER OF BRANCHING). ACTS AS AN ANTAGONIST OF EGF-R-MEDIATED
CC SIGNALING IN THE EYE (WHERE IT IS IMPORTANT FOR CELL
CC DETERMINATION) MIDLINE GLIA, CHOROTONAL ORGANS, WING AND OVARIAN
CC POLICIE CELLS.
-1- SUBUNIT: INTERACTS WITH DRK AND GAP1 PROTEIN OF THE RAS PATHWAY.
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE INNER SURFACE OF THE

```
CC PLASMA MEMBRANE.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN THE EMBRYONIC TRACHEAL SYSTEM,
CC DEVELOPING EEK IMAGINAL DISK, EMBRYONIC CHOROTONAL ORGAN
CC PRECURSORS, MIDLINE GLIA, WING DISK AND OVARY.
CC -I- DEVELOPMENTAL STAGE: FROM STAGE 7 OF OOGENESIS FOUND IN THE
CC POSTERIOR FOLLICLE CELLS AND DURING STAGE 9 WHEN THE FOLLICLE
CC CELLS MIGRATE POSTERIORLY OVER THE OOCTE NUCLEUS, EXPRESSION IS
CC SEEN IN THE DORSAL AND LATERAL CELLS, AND IS EXCLUDED FROM THE
CC VENTRAL CELLS. ONCE THE MIGRATION OF FOLLICLE CELLS IS COMPLETE
CC EXPRESSED IN THE DORSAL-ANTERIOR CORNER OF THE EGG CHAMBER.
CC -I- INDUCTION: BY THE BNL-RGF PATHWAY IN THE TRACHEAL SYSTEM AND BY
CC THE BEP RECEPTOR PATHWAY IN THE WING IMAGINAL DISK AND THE
CC FOLLICLE CELLS OF THE OVARY.
CC -I- DOMAIN: THE CYS-RICH DOMAIN IS RESPONSIBLE FOR THE LOCALIZATION C
CC THE PROTEIN TO THE PLASMA MEMBRANE.
CC -I- SIMILARITY: BELONGS TO THE SPROUTY FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstake
CC the European Bioinformatics institute. There are no restrictions on th
CC use by non-profit institutions as long as its content is in comm
CC modified and this statement is not removed. Usage by and for comme
CC entities requires a license agreement (See http://www.isb.ch/annc
or send an email to license@isb.sib.ch).
CC -----
CC EMBL; AF039842; AAC04257.1; -.
DR EMBL; AE003478; AAF4772.1; -.
DR FlyBase: FBgn0014388; sty.
KW Developmental protein; Membrane.
FT DOMAIN 104 135 SER-RICH.
FT DOMAIN 148 162 ASN-RICH.
FT DOMAIN 170 271 GLN-RICH.
FT DOMAIN 378 501 CYS-RICH.
FT DOMAIN 492 551 GLY-RICH.
FT DOMAIN 249 262 POLY-GLN.
FT DOMAIN 503 514 POLY-GLY.
FT CONFLICT 38 38 N -> T (IN REF. 1).
FT CONFLICT 262 262 Q -> QOO (IN REF. 1).
FT CONFLICT 312 312 P -> L (IN REF. 1).
SQ SEQUENCE 589 AA; 62388 MW; 7450B11987B17B20 CRC64;

Alignment Scores:
Pred. NO.: 1.35 Length: 589
Score: 110.00 Matches: 70
Percent Similarity: 32.73% Conservative: 39
Best Local Similarity: 21.02% Mismatches: 112
Query Match: 4.50% Indels: 112
DB: Gaps: 15

US-09-939-293-1 (1-1358) x SPY_DROME (1-589)
QY 518 GCACGAGAAGCGTCGATATCAACAACGTGCSCGACATGACGCCCTCTAACCSCCAGGAATCAC 5777
||| : : : : ||| : : : : ||| : : : : |||
Db 209 AlAtlrArgrseGlnInlSprAlcglvGlnInlAsprAnclgvlntPrHtPrHnIshts 2288
578 ATTCAAGTGTGAAACTCGAGTGAGGAGGTGCACACAGCTCTCCSGAAAAGAGAAC 6378
::: |||::: ::: ::: |||:::
Db 229 leuLeuleuleuProGlnlArAsnGlInhlslleutls ----- 2404
QY 638 AAGSTGCGAAGACACAGATGAAGAAGCTCCGTGAGAAAAACACAGAGAAAGGGAGAG 6977
||| : : : : ||| : : : : ||| : : : : |||
Db 241 ----- LeuGlInlInSlnglInlInhlslleugInlInglInglInglIn 254
698 CGGGGTGAGTGGAGAGAGAGAGGCTACCTGGCTGAGGATTGAGGGCTGAGAC ----- 7515
255 gInglInglInglInglInglInglInhlslleudInhlslInglInlAsnnglInhlslAlatrg 2744
752 -----ACTGSCCTGTCTCCCACTCAAGTGGGGAAGACAGAGGAGAT ----- 7933
Db 275 leuAlAtlrHtPrHnInglAlAtlrHSeValclvgeArsphtshlArsgluleuLeuhnlts 2944
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 794 -----GCCASCCTGGCCAGAGGTGTGCATGACTGTGTGTGACACCGAGA 8355
|||||
```

Db 295 SerHisLeuGlnAsnSerThrThrLysProProAlaSerLysGlnProAlaProProArg 314
 QY 836 AGAGCGGCGAGT----- 848
 Db 315 LeuGlyMetGlyLeuGlyLeuGlyLeuGlyLeuGlnProIleIleThrLys 334
 QY 849 ---CTGCGCCCGCCAAATCAGGCGAGACGCTTTGTGAGCTGTGAGTGCCTCTGTGGT 904
 Db 334 sglProThrProAlaThrGlnLysGlnArgMetHisAlaLeu-GlnGlnLeuLeuGlnP 354
 QY 905 CTCAGGCTGGCGTGCAGACCTCTTACGCCCTGGGACAGCA-----CCCTGTTTAA 958
 Db 354 roGlyGlyAlaGlyLeuGlnGlyGly---ProLeuValMetAlaGlyAspProSerLeu- 372
 QY 959 CATTTCACCCACCTCTG-----TACAGCTGCTCTTAC 991
 Db 373 --LeuAsnProIleValCysProArgCysGlyArgCysArgCysGlnGlnCysGlnSerP 392
 QY 992 CATTTCATTTTACCTCACACCCAAAGCATTTTGCCTACCTGGGTGACAGAGAGAGTCCTT 1051
 Db 392 roArgProLeuProGln-----ThrtProValCysAsnLysThrCysL 406
 QY 1052 TTTGTCATGCC-----CTTAGTTCAGACACTGTTTACCTGTTTACCTTATCTTAT 1102
 Db 406 euCysSerAlaGlnSerValIleAspTyrAlaSerCysLeuCysCysAlaLysAlaLeu- 425
 QY 1103 TACGTCGTCAAAATGATTAGTCTGTTGCTCTGTTGGATGCCAGTGTGGCAGGG 1162
 Db 426 -----PheTyrHisCysAlaArgAspAsnAspLeuAspLysAspAsp 440
 QY 1163 GGAGGGACCTGCTGCCAGT-----TGTACGATTTCTTTGAT- 1200
 Db 440 LysAsnGlyThrProCysValAspAsnProCysSerCysGlyProTyrLysArgThrGlnA 460
 QY 1201 -----GATTTCTGATG- 1212
 Db 460 rGTrpGlyTyrLeuGlyAlaLeuSerIlePheLeuProCysLeuTyrPheTyrTrpProM 480
 QY 1213 -----TGTTCCTGATCTGCCCCCCTGCTGCTGAGAGACAGCTGAGCCAAAGAGT 1264
 Db 480 eLarGlyGlyCysMetLysLeuGlyGlyLysCysTyrGlyArgPheAla----- 495
 QY 1265 GAAAACCTATTACTACTAGAGAGAGGGGTGACAGA 1299
 Db 496 -----GlyArgGlyCysArg 500
 RESULT 8
 TOP2_CABEL STANDARD; PRT; 1520 AA.
 ID TOP2_CABEL
 AC Q23670; Q27509;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Probable DNA topoisomerase II (EC 5.99.1.3).
 GN K12D12.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Swinburne J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Coles L.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining

CC of double-stranded DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z70213; CAA94177.1; -.
 DR EMBL; Z49069; CAA94177.1; JOINED.
 DR EMBL; Z49069; CAA88867.1; -.
 DR EMBL; Z70213; CAA88867.1; JOINED.
 DR HSP; P06786; 1BGW.
 DR WormPep; K12D12.1; CE06184.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR003957; CBFA_NFYB_topis.
 DR InterPro; IPR001241; DNA_topoisolI.
 DR InterPro; IPR002205; DNA_topoisolIV.
 DR Pfam; PF00204; DNA_gyraseb; 1.
 DR Pfam; PF00521; DNA_topoisolIV; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR PRINTS; PR00615; CCAATSUBUNITA.
 DR PRINTS; PR00418; PPI2FAMILY.
 DR ProDom; PD000616; DNA_topoisolI; 1.
 DR ProDom; PD000742; DNA_topoisolIV; 1.
 DR SMART; SM00433; TOP2c; 1.
 DR SMART; SM00434; TOP4c; 1.
 DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
 FT NP_BIND 196 201 ATP (POTENTIAL).
 FT ACT_SITE 840 840 DNA CLEAVAGE (BY SIMILARITY).
 FT SEQUENCE 1520 AA; 172333 MW; 016226697F41C360 CRC64;
 Alignment Scores:
 Pred. No.: 2.33 Length: 1520
 Score: 107.50 Matches: 70
 Percent Similarity: 37.42% Conservative: 49
 Best Local Similarity: 22.01% Mismatches: 110
 Query Match: 4.39% Indels: 89
 DB: 1 Caps: 15
 US-09-939-293-1 (1-1358) x TOP2_CABEL (1-1520)
 QY 92 CCTGTGTGGCTAAGCTTAAAGACCGGTGTTCTCAGATGATTAAGACCATGCGCACAA 151
 Db 987 ProValIleValAspTyrLysGln-----TyrHisThr 997
 QY 152 ACTGTGACGATGCGCTTGGATGTAACCTGTGCGGCTCTATTGACAGAAATACGAG 211
 Db 998 AspThrThrValLysPheValValLysLeuSerProGlyLysLeuArgGlnGlnArg 1017
 QY 212 CCTATTCCTTAAGTACTAGT-----GAACATGATGAGAGAGACAGTGTCTTTG 259
 Db 1018 GlyLysAspLeuHisGlnValPheLysLeuGlnAlaValIleAsnThrThrCysMetVal 1037
 QY 260 GTAACAGATAGACACCTTACCTTCTCTCAGACACATATGCGTTGATTGAAGTATT 319
 Db 1038 LeuPheAspAlaAlaGlyCysLeu-----ArgThrTyrThrSerProGlnAlaIle 1054
 QY 320 ACT-----GAATATCTAAG 334
 Db 1055 ThrGlnLysPheTyrAspSerArgGlnGlnLysTyrValGlnArgLysGlnTyrLeuLeu 1074
 QY 335 GCTGTTTATACCTTAACCTCT-----CTTACGACGACATATACAAAGTTACTGGGAAA 388
 Db 1075 GlyValLeuGlnAlaGlnSerLysArgLeuThrAsnGlnAlaArgPheIleLeuAlaLys 1094
 QY 389 ATGAATTCAGAG-----GAGAGATGAAGTGTGCGACAGTGATCATATA 430

[illegible]

```

RESULT 9
TEGU_HSV11 STANDARD: PRT: 3164 AA.
AC P10220;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein (Virtion protein UL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RL herpes simplex virus type 1."
J. Gen. Virol. 69:1531-1574(1988).
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BPLF1, HVS-1 64, VVY 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14112:CAA32311.1; -.

```

DR PIR: I30085; WMBEH6.
DR InterPro: IPR005210; Herpes_UL36.
DR Pfam: PF03586; Herpes_UL36; 1.
FT REPEAT.
KW DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q
SQ SEQUENCE 3164 AA: 335857 MW: CC5D11FFAF9FE334 CRC64:

Alignment Scores:

Pred. No.:	3.04	Length:	3164
Score:	106.50	Matches:	85
Percent Similarity:	33.12%	Conservative:	21
Best Local Similarity:	2.55%	Mismatches:	129
Query Match:	4.41%	Indels:	85
DB:	1	Gaps:	17

US-09-939-293-1 (1-1358) x TEGU_HSV11 (1-3164)

1252 CAGCTGTCCCTCACAGACAGTGGGGCAGATCA-----GAG 1217

DB 2381 GULEUAIASERASpaspserIGYIProIAProLeuLeuAIAProAsProLeuSer 2600

[illegible]

11EC CCCCCC
CCCCCCCCCCCCCCCCCC

```

Db      2620  aValThrAaArgG|uAlAArgProG|vAlProA|G|uSerThrArArProA|a----- 2638
          |||      |||::|||  |||  |||

```

QY 1114 TTGACGACGTAATAAGACTGAAACAGGTTAAACAGTTGCTGAACCTTAAGGCATGAC 1055

Db 2639 -----ProValGlyProArgaspasphear 2647

1054 AAAA--AGGACTCTCTCTGACCCAGGTAGGCCAAATGCTTGGGTGTGAGGTA AAA 998

Db 264 / gargleuProserProginserProAlaproAspalathralaProargProPr 266 /

27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054

[illegible]

dh 3678 rG]vserArqa]aArqArqHisaArqArqa]aArqSerLeuA]aArqa]aThrg]na]ase 3698

877 AGGCGTCTCGCCTGATTGGCCAGGGCAGGACCTGCCGCCCTCTTCTCGGTGCACAGACAGT 818

Db 2698 rAlaThrThrGlnGlyTrp-----ArgProProAlaLeu---ProAspThrVa 2713

817 CATGCCAACCCCTGGGCAGGGTG---GCATCTGCCCCCTGCTTCCCACTGAGTGGGGAGA 761

Db 2/13 lAlaProValThrAspPheAlaArgProProAlaProProLysProProGluProAlaPr 2/33

[illegible][illegible]

```

db      2753  erpProA]atEubProT]easbProVaj]ProBProProVaj]A]atHrc]v-----T 2769
          ::|||      |||      |||:::|||||      ::      |||

```

QY 652 TGCTTCTGCCAGCTTGGTTT-----CTGCTTTCGGGAGAGCTGGTGCAC 608

Db 2769 hrvalLeuProGlyGlyGluAsnArgArgProProLeuThrSerGly-----ProAlaP 2787

607 CTCTTCCACCTGCAGTTTCACCAGCTGAATGTGATTCTTGGGGTTATAGAGGCTGATC 548

Db 2787 rothrproproargval---provalglyglyproglinargargleuthrargpro---- 2804

xy 34 / 1666CAG11GAI1GCGC11C16C16CCA1C1GAAAGACCAAC16CAG1CA--- 492

```

QY 491 --TCCAGTGGTTCCAGCTTCAGTCTGTGTTG----- 453
DB 2817 roserProtrPaspProAlaPaspProThrAlaPaspProValLeuGlyArgAsnProAlaGluP 2837
QY 452 --AAGTCAATCAGCTGCGCTCCATGATCAGCCAGCCACACTTATCTCTCT 399
DB 2837 rothrSerSer-----SerProAlaGlyProSerProProPro 2849

RESULT 10
WASP_MOUSE
ID WASP_MOUSE STANDARD: PRT: 520 AA.
AC P70315;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Wiskott-Aldrich syndrome protein homolog (WASP).
GN WAS OR WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=96115600; PubMed=8666397;
RA Derry J.M.J., Miedemann P., Blair P., Wang Y., Kerns J.A.,
RA Lemahieu V., Godfrey V.L., Wilkinson J.E., Francke U.;
RT "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)
RT gene is highly conserved and maps near the scurfy (sfi) mutation on
RL the X chromosome.";
RL Genomics 29:471-477(1995).
CC -1- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC -1- MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION
CC (BY SIMILARITY)
CC -1- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -1- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U54788; AAC52556.1; -.
DR MGI; MGI:105059; Was.
DR InterPro: IPR0000095; PAKbox/RhoGndng.
DR InterPro: IPR0000697; RANBP1_WASP.
DR InterPro: IPR001960; WH1.
DR InterPro: IPR003124; WH2.
DR Pfam; PF00568; WH1; 1.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00461; WH1; 1.
DR SMART; SM00246; WH2; 1.
DR PROSITE; PS50108; CRIB; 1.
KW Repeat.
FT DOMAIN 41 147 WH1.
FT DOMAIN 240 253 CRIB.
FT REPEAT 354 363 GRSGPLPPXP MOTIF 1.
FT REPEAT 393 402 GRSGPLPPXP MOTIF 2.
FT DOMAIN 162 167 POLY-PRO.
FT DOMAIN 314 321 POLY-PRO.
FT DOMAIN 324 341 POLY-GLY.
FT DOMAIN 368 373 POLY-PRO.
FT DOMAIN 376 379 POLY-PRO.
FT DOMAIN 384 390 POLY-PRO.
FT DOMAIN 397 403 POLY-PRO.
FT DOMAIN 408 424 POLY-PRO.

```

```

FT DOMAIN 503 520 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 520 AA; 54191 MW; 9C223733C59F0C8A CRC64;

Alignment Scores:
Pred. NO.: 2.6 Length: 520
Score: 106.00 Matches: 37
Percent Similarity: 44.44% Conservative: 33
Best Local Similarity: 41.11% Mismatches: 3
Query Match: 4.39% Indels: 17
DB: 1 Gaps: 2

US-09-939-293-1 (1-1358) x WASP_MOUSE (1-520)
QY 808 CTTGGGAGAGGTGGATGCTGCTTCCACATGATGGGAGACAGGAGTGTG 749
DB 379 ProGlyArgGlyGlyProProProProProProAlaThrGlyArgSerGlyProPro 398
QY 748 CTCAGGCCCTCAATCCATCAGCAGTAGGCTCTGCTCGCATCTAGCCGCTCTCC 689
DB 399 ProProProLeu--ProGlyAlaGlyProProAlaProProProProProPro 418
QY 688 TTCTCTCTGTGTTTCTACGAGAGCTCTTATCTGTCTCTGCGAGCTTGTTCTGC 629
DB 418 roProProProProCysProGly----- 425
QY 628 TTTCCGGAGAGCTGTCACCTCTTCCACCTGCA---GTTTACACAGCTGAATGTGAT 572
DB 426 --SerGlyProAlaProProProProProProProProValSerGlyGly--SerProAla 444
QY 571 CTTGGCGGTTATAGAGGCGCTGATC 548
DB 445 ProGlyGlyGlyArgGlyAlaLeu 452

RESULT 11
JAG1_RAT
ID JAG1_RAT STANDARD: PRT: 1219 AA.
AC Q63722; P70640;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Jagged 1 precursor (Jagged).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=95211842; PubMed=7697721;
RA Lindsay C.E., Shawber C.J., Boulter J., Weinmaster G.;
RT "Jagged: a mammalian ligand that activates Notch1.";
RL Cell 80:909-917(1995).
CC -1- FUNCTION: Ligand for multiple Notch receptors and involved in the
CC mediation of Notch signaling. May be involved in cell-fate
CC decisions during hematopoiesis. Enhances fibroblast growth
CC factor-induced angiogenesis (in vitro). Seems to be involved in
CC early and late stages of mammalian cardiovascular development.
CC Inhibits myoblast differentiation. May regulate fibroblast
CC growth factor-induced angiogenesis.
CC -1- SUBUNIT: Interacts with NOTCH1, NOTCH2 and NOTCH3 (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Widely expressed in a variety of tissues.
CC -1- DEVELOPMENTAL STAGE: Expression is seen in E11.5-E14.5 embryos in
CC four distinct regions of the ventricular zone in the developing
CC spinal cord.
CC -1- SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DSL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L38483; AAB06509.1; .
 DR HSSP: P00740; 1EDM.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001774; DSL.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF01414; DSL; 1.
 DR Pfam: PF00008; EGF; 14.
 DR PRINTS: PR00010; EGFBLD.
 DR SMART: SM00051; DSL; 1.
 DR SMART: SM00179; EGF_CA; 10.
 DR SMART: SM00001; EGF_like; 6.
 DR SMART: SM00214; VMC; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 10.
 DR PROSITE: PS00022; EGF_1; 16.
 DR PROSITE: PS01186; EGF_2; 12.
 DR PROSITE: PS01187; EGF_Ca; 8.
 DR PROSITE: PS01187; EGF_Ca; 8.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Developmental protein;
 KW Repeat; Transmembrane; Signal.
 FT SIGNAL 1 33 POTENTIAL.
 FT CHAIN 1 1219 JAGGED 1.
 FT DOMAIN 34 1067 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1068 1093 POTENTIAL.
 FT DOMAIN 1094 1219 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 167 229 DSL.
 FT DOMAIN 230 266 EGF-LIKE 1 (ATYPICAL).
 FT DOMAIN 296 334 EGF-LIKE 2.
 FT DOMAIN 336 372 EGF-LIKE 3.
 FT DOMAIN 374 410 EGF-LIKE 4.
 FT DOMAIN 412 448 EGF-LIKE 5.
 FT DOMAIN 450 485 EGF-LIKE 6.
 FT DOMAIN 487 523 EGF-LIKE 7.
 FT DOMAIN 525 561 EGF-LIKE 8.
 FT DOMAIN 561 627 EGF-LIKE 9.
 FT DOMAIN 629 665 EGF-LIKE 10.
 FT DOMAIN 667 703 EGF-LIKE 11.
 FT DOMAIN 705 741 EGF-LIKE 12.
 FT DOMAIN 744 780 EGF-LIKE 13.
 FT DOMAIN 782 818 EGF-LIKE 14.
 FT DOMAIN 856 890 EGF-LIKE 15.
 FT DISULFID 300 312 BY SIMILARITY.
 FT DISULFID 306 322 BY SIMILARITY.
 FT DISULFID 324 333 BY SIMILARITY.
 FT DISULFID 340 351 BY SIMILARITY.
 FT DISULFID 345 360 BY SIMILARITY.
 FT DISULFID 362 371 BY SIMILARITY.
 FT DISULFID 378 389 BY SIMILARITY.
 FT DISULFID 383 398 BY SIMILARITY.
 FT DISULFID 400 409 BY SIMILARITY.
 FT DISULFID 416 427 BY SIMILARITY.
 FT DISULFID 421 436 BY SIMILARITY.
 FT DISULFID 438 447 BY SIMILARITY.
 FT DISULFID 454 464 BY SIMILARITY.
 FT DISULFID 458 473 BY SIMILARITY.
 FT DISULFID 475 484 BY SIMILARITY.
 FT DISULFID 491 502 BY SIMILARITY.
 FT DISULFID 496 511 BY SIMILARITY.
 FT DISULFID 513 522 BY SIMILARITY.
 FT DISULFID 529 540 BY SIMILARITY.
 FT DISULFID 534 549 BY SIMILARITY.
 FT DISULFID 551 560 BY SIMILARITY.
 FT DISULFID 578 605 BY SIMILARITY.
 FT DISULFID 599 615 BY SIMILARITY.
 FT DISULFID 617 626 BY SIMILARITY.
 FT DISULFID 633 644 BY SIMILARITY.
 FT DISULFID 638 653 BY SIMILARITY.
 FT DISULFID 655 664 BY SIMILARITY.

FT DISULFID 671 682 BY SIMILARITY.
 FT DISULFID 676 691 BY SIMILARITY.
 FT DISULFID 693 702 BY SIMILARITY.
 FT DISULFID 709 720 BY SIMILARITY.
 FT DISULFID 714 729 BY SIMILARITY.
 FT DISULFID 731 740 BY SIMILARITY.
 FT DISULFID 748 759 BY SIMILARITY.
 FT DISULFID 753 768 BY SIMILARITY.
 FT DISULFID 770 779 BY SIMILARITY.
 FT DISULFID 786 797 BY SIMILARITY.
 FT DISULFID 791 806 BY SIMILARITY.
 FT DISULFID 808 817 BY SIMILARITY.
 FT DISULFID 824 835 BY SIMILARITY.
 FT DISULFID 829 844 BY SIMILARITY.
 FT DISULFID 846 855 BY SIMILARITY.
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 559 559 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1064 1064 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1219 AA; 134325 MW; 65D4CFC238A0E204 CRC64;
 Alignment Scores:
 Pred. No.: 3.75 Length: 1219
 Score: 104.50 Matches: 71
 Percent Similarity: 29.67% Conservative: 37
 Best Local Similarity: 19.51% Mismatches: 131
 Query Match: 4.27% Indels: 125
 DB: 1 Gaps: 19
 US-09-939-293-1 (1-1358) x JAG1_RAT (1-1219)
 QY 481 AACCACTGGATGACGCTGGTCT----- 507
 Db 666 AsnAsnIleAsnAspCysSerGlnAsnProCysHisTyrGlyGlyThrCysArgSerLeu 685
 QY 508 TTCAGAGTGGCAGCAGAGAGCTGCATATCAAACTGGCGAGATCAGGCTGTATACCGC 567
 Db 686 ValAsnAspPheTyrCysAspCysLysAsnGlyTyrPheGlyThrCysHisSerArg 705
 QY 568 CAGGAATCATTCACCTGCTGAAACTGCAGGTGGAAGGGTGCACCACTCTCCGGAA 627
 Db 706 AspSer-----GlnCysAspGluAlaThrCysAsnAsnGlyGlyThrCysTyrArgGlu 723
 QY 628 -----AGCAAAACCAAGCT 642
 Db 724 ValAspThrPheLysCysMetCysProGlyGlyTyrPheGlyGlyThrCysAsnIleAla 743
 QY 643 GGCAGAGCAGACAGATAGAAGACTCCGTGAGAAACACAGAGGAGGAGAGCGGC 702
 Db 744 ArgAsnSerSerCysLeuProAsnProCysHisAsnGlyGly----- 757
 QY 703 TGAAGTCGGAGCAGAGGCTACCTGCGTGAAGATTGAGGGCTGAGACACAGCTGCTGC 762
 Db 758 -----ThrCysValVal 761
 QY 763 TCCCCACAGTGGGGAAGCAGGGGACAGATGCACCCCTGCCAGGGTGGCATGACTGT 822
 Db 762 -----AsnGlyAspSer---PheThrCysValCysLysGluGlyTyrPheGluGlyPro 777
 QY 823 CTGTGCACCCGAGAGAGCGGCGAGGTCTGCCCTGGCCA----- 861
 Db 778 IleCysThrGlnAsnThrAsnAspCysSerProHisProCysTyrAsnSerGlyThrCys 797
 QY 862 ATCAGCGGAGAGCCCTTTGTGACSTGTGAGTGCCTCTGTGCTCAGGCTGGCCGTGGA 921
 Db 798 ValAspGlyAspAsnTyrPheArgCysGluCys-----AlaProGlyPheAlaGly 814
 QY 922 CCT-----GGTTCTTAGCCCTTGGGCACTGCACCCCTGTTTAACATTACCCG 969

GN LAMB2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91106216; PubMed=2102682;
 RA Hoeger T.H., Zatloukal K., Weizenegger I., Krohne G.;
 RT "Characterization of a second highly conserved B-type lamin present
 in cells previously thought to contain only a single B-type lamin.";
 RL Chromosoma 99:379-390(1990).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=9139548; PubMed=2102440;
 RA Hoeger T.H., Zatloukal K., Weizenegger I., Krohne G.;
 RL Chromosoma 100:67-69(1990).
 CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
 LAYER ON THE NUCLEOPASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
 WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
 AND MAY ALSO INTERACT WITH CHROMATIN.
 CC -1- SUBCELLULAR LOCATION: NUCLEOPASMIC SIDE OF THE INNER NUCLEAR
 MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: B2 (shown here) and B3 (AC
 p4680); may be produced by alternative splicing.
 CC -1- PTM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS
 FARNESYLATION AND PHOSPHORYLATION. INCREASED PHOSPHORYLATION OF
 THE LAMINS OCCURS BEFORE ENVELOPE DISINTEGRATION AND PROBABLY
 PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.
 CC -1- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY
 CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISTINGUISHED AND
 FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TETROPHASE,
 RESPECTIVELY.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC THIS IS A B TYPE LAMIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54098; CAA38032.1; -
 DR PIR: S21609; S21609.
 DR PIR: B48315; B48315.
 DR MGP: MGI:96796; Lmbd2.
 DR InterPro: IPR001664; IF.
 DR InterPro: IPR001322; IF_tail.
 DR InterPro: IPR001230; Pfenyl_site.
 DR Pfam: PF00038; filament_1.
 DR Pfam: PF00932; IF_tail; 1.
 DR PROSITE: PS00226; IF: 1.
 KW Intermediate filament; Coiled coil; Nuclear protein; Lipoprotein;
 KW Prenylation; Phosphorylation; Alternative splicing.
 FT DOMAIN 1
 FT 1 26 HEAD.
 FT DOMAIN 2
 FT 27 378 ROD.
 FT DOMAIN 3
 FT 379 592 TAIL.
 FT DOMAIN 4
 FT 62 73 COIL 1A.
 FT DOMAIN 5
 FT 74 73 LINKER 1.
 FT DOMAIN 6
 FT 207 73 COIL 1B.
 FT DOMAIN 7
 FT 208 234 LINKER 2.
 FT DOMAIN 8
 FT 235 378 COIL 2.
 FT DOMAIN 9
 FT 414 419 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 10
 FT 561 574 ASP/GLU-RICH (HIGHLY ACIDIC, COULD BE
 INVOLVED IN CHROMATIN BINDING).
 FT LIPID
 FT 589 589 FARNESYL (BY SIMILARITY).
 FT SEQUENCE 592 AA; 67029 MW; 7D5AC51BC9A5041E CRC64;
 Alignment Scores:
 Pred. No.: 4.39 Length: 592
 Score: 103.00 Matches: 63

Percent Similarity: 41.44% Conservative: 58
 Best Local Similarity: 21.58% Mismatches: 105
 Query Match: 4.21% Indels: 66
 DB: 1 Gaps: 13
 US-09-939-293-1 (1-1358) x LAMB2_MOUSE (1-592)
 QY 194 ATTCGACAGAAATAGAGACCTCATCCCTTAGTAGTGAAGCATGTGAGAGACAGT 253
 DB ::::::::::: ||::::::::: :::: ||:::
 QY 162 LeuAlaLysAlaGluAspGlyHisAlaValAlaLysGlnLeuGluGluThrLeu 181
 QY 254 TCTTTGGTAAAGATAGACACCTTACCTTCTCTCTCAGACACATATGCGTGAATGAA 313
 DB :::: ||:::
 QY 182 MetArgValAsp-----LeuGlnAsnArgCysGlnSerLeuGlnGlu 195
 QY 314 GCTATTACTGATATATACCTAAGCTGTTAT-----ACCTTAATCTCTTTACCGACA 367
 DB :::: ||::::::::: ||:::
 QY 196 GluLeuAla---PheSerLysSerValPheGlnGluGluValArgGluThrArgArgArg 214
 QY 368 TATCAAGATTACTTGGGAAATGATATGATTCAGAGAGGAAAGATGAAGTGGCAGGTGATC 427
 DB :::: ||:::
 QY 215 HisGluArgArgLeuValGluValAspSerSerArgGlnGlnGluThrAspPheLysMet 234
 QY 428 ATAGAGCCAGAGCTGAGATGATCTTCAAAACACCAAGCTACTTGAG----- 475
 DB :::: ||:::
 QY 235 AlaGlnAlaLeuGluAspLeuArgSerGlnHisAspGlnGluValArgLeuThrArgVal 254
 QY 476 --CTGGAACCACTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 532
 DB :::: ||:::
 QY 225 GluLeuGlnGlnThrThrTyGlnAlaLysLeuAspAsnAlaLysLeuSerAspGln--- 273
 QY 533 TATCAAACTGGCGCAGATCAGGACCTCTTAAACCGCAGGAATCAATTCAGCTGTGAA 592
 DB :::: ||:::
 QY 274 -----AsnAspLysAlaAlaHisAlaValArgGluGluLeuLysGluValArg 289
 QY 593 CTGCAGGTGGAAGAGGTG-----CACCACTCTCCCGAAAGCAAGAACCAAGCTGCA 646
 DB :::: ||:::
 QY 290 MetArgValGlnSerLeuSerTyGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 309
 QY 647 GAGACAGCATGAGAAGAGCTCCGTCAG----- 673
 DB :::: ||:::
 QY 310 GluAsnHisLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 329
 QY 674 -----AAACACAG 718
 DB :::: ||:::
 QY 330 MetLeuAspAlaLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 349
 QY 719 -----GCTAC 724
 QY 350 AlaGluThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 369
 QY 725 CTGCGTGGAGATGAGAGG-----CTGAGACACACTGCGCTGCTCCGCAC----- 769
 QY 370 ArgLysLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 389
 QY 770 --TCAGTGGGGAAG 826
 DB :::: ||:::
 QY 827 GCACCGAAGAAAG 880
 DB :::: ||:::
 QY 409 --GlyGlnArgGlyGlyLysArgArgArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 426
 QY 881 TGAGCTGTAGTGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 916
 DB :::: ||:::
 QY 427 ArgAlaSerArgValSerSerGlySerArgLeuAla 438
 RESULT 14
 REST_CHICK STANDARD: PRT: 1433 AA.
 AC 042184; 042228; 057563; 057564;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)


```
OY 248 GCAGTGTCTTTG-----GTAACAGATAGCACCCT-----ACC 280
      :: ||| ||| |||::|||
Db 1377 ArgLeuProLeuHeuValLysValArgAspThrThrGlnGluProCysGlyArgLeuSer 1396
      :::::
OY 281 TTTCTCTCTCAG-----ACCACATATGCGTGTATT--GAAGCTATT-----ACTGAA 325
      |||::: ::::: |||::: ::::: |||:::
Db 1397 PheMetLysGluProLysSerThrArgGlyLeuValHisGlnAlaIleCysAsnLeuAsn 1416
      :::::
OY 326 TATACCTAGAGCTGTTTACTTACTTCTTACTTCTTACCGACATATACAGTTTACTTGGG 385
      ||| ::::: ||| ||| |||
Db 1417 IleThrLeuProIleThrLysGluSerGluSerAspGlnGlnGluIle 1436
      |||
OY 386 AAATGAAATTCAGAGGAGGAAGATGAA----- 412
      ||| ||| |||::: ||| |||
Db 1437 AspMetThrSerLeuLysAsnAspGluThrGluSerThrGluThrSerValLeuLysSer 1456
      ||| ::::: ||| ||| |||
OY 413 -----GTGTGGCAGGTGATCATAGAGCCAGAGCTGAGATGACTTCMAAACACAGAG 466
      ||| ::::: ||| ||| ::::: |||
Db 1457 HisLeuValasnGluValProValLeuAlaSerProAspLeuSerGluValSerGlu 1476
      ||| ::::: ||| ||| |||
OY 467 TACTTGAGCTGGAACCACTTGATGACTGCAGTGTGCTTCAGAGATGGCA----- 520
      ::::: ||| ||| ||| |||
Db 1477 ---MetLysGlnAspLeuIleLysMetThrAlaIleLeuThrThrAspValSerAspLys 1495
      ::::: ||| ||| ||| |||
OY 521 -----GCAGAAGCTGCATATCAAACTGGCGGCAGATCAG 553
      ::::: ||| ||| ||| |||
Db 1496 AlaGlySerIleLysValLysGluLeuValLysAlaAlaGluGluProGlyGluPro 1515
      ||| ||| ||| ||| |||
OY 554 GCCTCTATTAACCGCCGGAATTCACATTACGCTGCTGAACCTGCAGTGGAAGAGTGCAC 613
      ||| ||| ||| ||| |||
Db 1516 PheGluIleValGluArg-----ValLysGlnAspLeuGluLysValAsn 1530
      ||| ||| ||| ||| |||
OY 614 CAGCTCTCCCGAAGCAAGCAAAACCAAGCTGGCAGACACAGATAGAAGAGCTCCGTCAG 673
      ::::: ||| ||| ||| |||
Db 1531 GluIleLeuArgSerGlyThrCysThrArgAspGluSerSerValGlnSerSerArgSer 1550
      ::::: ||| ||| ||| |||
OY 674 AAA-----ACACAGAGAGAA-----GGGAGAGAGCGGGCTGAGTGC 709
      ::::: ||| ||| ||| |||
Db 1551 GluArgGlyLeuValGluGluGluThrPvalIleValSerAspGluIleGluGluAla 1570
      ::::: ||| ||| ||| |||
OY 710 GAGCAGAGGCGCTACCTGCGCTGAGAGATTGAGGGCT--GAGCACACTGCCCTGCTCC 766
      |||::: ||| ||| ||| |||
Db 1571 ArgGlnLysAlaProLeuGluIleThrGluLutyrProCysValGluValArgIleAspLys 1590
      |||::: ||| ||| ||| |||
OY 767 CACTCAAGTGGGGAAGAGAGGCGCAGATGCACCTGCCCAAGGTTGGC----- 814
      ||| ||| ||| ||| |||
Db 1591 GluIleLysGlyLysValGluLysAspSerThr-----GlyLeuValAsnTyr 1606
      ||| ||| ||| ||| |||
OY 815 -----ATGACTGTCTGTGCACG----- 832
      ||| ||| ||| ||| |||
Db 1607 LeuThrAspAspLeuAsnThrCysValProLeuProLysGluGlnLeuGlnThrValGln 1626
      ::::: ||| ||| ||| |||
OY 833 AGAAGAGCGCGCAGGCTCGCCCT-----GGCCAATCAGGCGAG 871
      ::::: ||| ||| ||| |||
Db 1627 AspLysAlaGlyLysLysCysGluAlaLeuAlaValGlyArgSerSerGlu 1643
      ::::: ||| ||| ||| |||
```

Search completed: February 20, 2003, 14:58:18
Job time : 47.5 secs

THIS PAGE BLANK (USPTO)

Query	Match:	25.06%	Indels:	2
DB:	11	Gaps:	1	
US-09-939-293-1 (1-1358) x Q8RI08 (1-157)				
QY	20	ATGCGCGGTCTGAGAGTATGGCTGTCCGCGAGCGTAACTTCAATCTTCAGGTACAGAC	79	
DB	1	MetalaalaleuagsetrtpvalthrargserValcysserleuphenrtyrargln	20	
QY	80	TGTTGTGTCTTCCTGTGTGGCTTAACCTTAAAGACGGGTCTTCAGAAATGATAGGA	139	
DB	21	ArpHe-----ProValleuAlaanserIysrIysrArgcysPheSerGluIleuIels	38	
QY	140	CCATGGCAAAACTGTGACGATTGGCTTTGGAGTAACCTGTGTGGCTTCTATTGCA	199	
DB	39	ProtrphistIysrthValleuthrGlyPheGlymethrIleucysAlaValproIleAla	58	
QY	200	CAGAAATCAGAGCCTCATCTCCCTTAGTAGTGAACATTGATGAGAGACAGTCTTTG	255	
DB	59	GlnIysSerGluProGlnSerIleuSerAsnIuAlaIleuMetArgAlaValSerIeu	78	
QY	260	GTACAGATTACACCTTACTCTTCCTGTACAGCCATATGGTGTGATGAAGCATTT	319	
DB	79	ValthrIspSerthrSerthrPheIeuSerIthrThrtyrAlaIleuIleGluAlaIle	98	
QY	320	ACTGAATATTAAGGCTGTATTATACCTTAATCTTCCTTTACGACAAATATACAGTTTA	379	
DB	99	ThrIuIyrrhrIyAlaValIyrrhrIleuValSerIeuIyrrArgIuIyrrhrSerIeu	118	
QY	380	CTTGGGAAATGAATTCACAGAGGAGGAAGATGACAGTGGCAGGTGATCATAGACCGA	439	
DB	119	LeuGlyIysMetAaSerGlnGluAspIuValIyrrGlnValIleIleGlyAlaArg	138	
QY	440	GCTGAGTACT 451		
DB	139	ValGluValSer 142		
RESULT 2				
Q9P682	PRELIMINARY;	PRT;	803 AA.	
AC	Q9P682;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-DEC-2001 (TREMBlrel. 19, last sequence update)			
DT	01-JUN-2002 (TREMBlrel. 21, last annotation update)			
DE	Related to PCF11 component of pre-mRNA 3'-end processing factor CP			
DE	I			
GN	B1D1.390.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Sordariales; Sordariaceae; Neurospora.			
OX	NCBI_TaxID=5141;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Schulze U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,			
RA	Nyakatura G., Mewes H.W., Mannhaupt G.;			
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
RP	[2]			
RP	SEQUENCE FROM N.A.			
RA	German Neurospora genome project;			
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL355927; CAB91288.2; -			
DR	InterPro: IPR002965; P-rich_extensn.			
DR	PRINTS; PR01217; PRICHEXTENS.			
Q0	SEQUENCE 803 AA; 84196 MW; F9955AEC3BC1E5B CRC64;			

Alignment Scores:	
Pred. No.:	0.0769
Score:	123.00
Percent Similarity:	42.41%
Best Local Similarity:	31.01%
Query Match:	5.10%
DB:	3
Length:	8039
Matches:	49
Conservative:	18
Mismatches:	72
Indels:	19
Gaps:	6

[illegible]

Alignment Scores:	
Pred. No.:	0.129
Score:	120.00
Percent Similarity:	39.58%
Best Local Similarity:	21.91%
Query Match:	4.91%
DB:	4
Gaps:	11
Length:	600
Matches:	62
Conservative:	50
Mismatches:	87
Indels:	84
Gaps:	11

US-09-939-293-1 (1-1358) x Q9NSB2 (1-600)

```

OY 251 GGTCTTGTGTAAGATGACACCTTACCTTCTCTCTCAGACCAATAT---GGCTTG 307
DB 268 VALALALEuLysLysAspValAspAlaAlaPheMetAsnLysSerAspLeuGluAlaAsn 287
OY 308 ATTTGAAGCTATTACT-----GAATATACTAAAGCGTGTAT-----ACC 346
DB 288 VALAspThrLeuThrClnGlnLysAspPheLeuLysThrLeuTyrtetGluGln 307
OY 347 TTTACTTCTCTTTACCGCAATATACACTTTACTTGGAAAAATGAAATTCAGAGAGAA 406
DB 308 LeuLeuGlnSerHisLieserGluThrSerValLleValLysMetAspSerAsp 327
OY 407 GATGAAGTGTGGAGGTGATAGAGAGCCAGAGCCAGATGATGATCAAAACCAAGAG 466
DB 328 LeuAsnLeuAspLylLlele-----AlaGluValLysLagLntyrGluGln 343
OY 467 TACTTGAACTGGAAACCACTTGATGACTGAGTTGTCTTTCAAGAGATGGACAGAA 526
DB 344 VALAlaArg-----ArgSerArgAlaAspAlaGlu 353
OY 527 GCTGCATATCAAACTGGCCAGATCAGGCTCTATAACCCGCAAGATACATTCAGCTG 586
DB 354 AlATPTyGlnThrLysTyrgLumetGlnValThrAlaGlyLnhLysAspAsn 373
OY 587 GTGAACCTCAGGTGAGAGAGGCGACCACTCTCCCG-----AAAGA 631
DB 374 LeuAlaAsnLleArgAsnLulLleAsnGluLeuThrArgLeuLleGlnArgLeuLys 393
OY 632 GAACCAACCTGCGAGAACACAG-----ATAGAAGAGCTCCGTCGAAACACAG 682
DB 394 GluLleGlnHisLalLysLalGlnArgLalLysLeuGlnAlaLalValAlaGlu 413
OY 683 GAGGAAGGGAG-----694
DB 414 GlnGlnGlyLalThrLeuSerAspAlalLysCysLysLeuAlaLaspLeuGluCysAla 433
OY 695 ----GAGCGGGCTGAGTCGAGACAGAGAGGCTTACCTGCGAGAGATTGAGGGCTGAGCAC 751
DB 434 LeuGlnGlnAlalLysGlnAspMetAlaArgGlnLeuArg-----GluTyr 448
OY 752 ACTGCCCTGTCTCCCACTCAGTGGGGGAAAGCAGGAGATGCCACCTGCCAGAGGTT 811
DB 449 GlnGluLeuMetAsnAlalLysLeuGlyLeuAspLleGluLleAlaThrTyrArgArgLeu 468
OY 811 -----811
DB 469 LeuGluGlyLglLysSerArgLeuCysGluValGlyProValAsnLieserValSer 488
OY 812 -----GGCATGACTCTCTGTGCACCGACA-----835
DB 489 SerSerArgGlyLglLeuValCysGlyProGluProLeuValAlaLysThrLeuSer 508
OY 836 AGAGCGGAGAGGCTGCGCCCTGCAATCAGAGGAGAGCCCTTGTGACCTGAGTGGCC 895
DB 509 ArgGlyGlyValThrPheSerGlySerSerValCysAlaThrSerGlyValLeuAla 528
OY 896 TCCTGTGTGT 904
DB 529 SerCysGly 531
RESULT 4
O95YL3 PRELIMINARY: PRT: 422 AA.
AC O95YL3;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE CS-OTX protein.
GN CS-OTX.
OS Clona savigny1.

```

OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 OC Phlebobranchia; Clonidae; Clona.
 OX NCBI_TaxID=51511;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21450572; PubMed=11566860;
 RA Satou Y., Imai K.S., Satoh N.;
 RT "Early embryonic expression of a LIM-homeobox gene Cs-1hx3 is
 RT downstream of beta-catenin and responsible for the endoderm
 RT differentiation in Clona savigny embryos";
 RL Development 128:359-3570(2001).
 DR EMBL; AB057732; BAB68341.1; -;
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR Prodom: PD00010; Homeobox; 1.
 DR PROSITE: PS00027; HOMEBOX_1; UNKNOWN_1.
 DR PROSITE: PS0071; HOMEBOX_2; 1.
 SQ SEQUENCE 422 AA; 44376 MW; 617848FP97CDC896 CRC64;

Alignment Scores:
 Pred. No.: 0.177 Length: 422
 Score: 118.00 Matches: 82
 Percent Similarity: 33.53% Conservative: 30
 Best Local Similarity: 24.55% Mismatches: 105
 Query Match: 4.89% Indels: 117
 DB: 5 Gaps: 16

US-09-939-293-1 (1-1358) x Q95YL3 (1-422)

```

OY 952 AGGTGACAGTGCSSAAGGCTTAAGAACAGATCCAGCCGAAAGCTTGAGACAGAGAGC 893
DB 117 ArgAlaLysCysArgGlnGlnValGlnGlnGlnGlnLysSerGly 136
OY 892 ACTCACAGCTCACAAAGGCTCTCGCTGATGGCCAGGAGAGACCTGCCCTTTCT 833
DB 137 SerGlySerSerSer-----AlaAsnAsnSerGlySerSer 151
OY 832 CGGTGCAGAGACATGCAACCTGGGAGAGGAGCA-----TCCTCCC 785
DB 152 GlySerThrAsnSerAsnAsnAsnSerSerSer-AlaSerLysLysSerPro 171
OY 784 TCGTCTCCCACTGAGTGGGAGACAGAGCAGTGTCT-----CAGAGCC 740
DB 171 ornThrProThrProThrAla-ThrClyProAlaLaserThrSerProPheGln 191
OY 739 TCAATCCTACAGGATGAGCTCTGCTCCAGCTCAGCCGCTCCCTCTCTCTCTCT 684
DB 191 eutThrAlaProProAlaHisGlnProLeuAlaGlnSerSerSerProProLeuPro 211
OY 683 ----CTGTGTTTTGACGAGAGCTCTGATCTGTCTGTCAGAGCTGTGTTCTCTCT 626
DB 211 erProValSerSerAsnGlnLleGlyProSerValThrProSer-----GluLeu 229
OY 625 CCGAGAGAGCTGGTGCACCTT-----603
DB 229 ecLysSerGlyAlaProLeuAsnHisGlnLysSerHisGlyAlaHisSerGlyMet 249
OY 602 -----CSACCTGCAGTTTACCA-----585
DB 249 alHisThrGlySerSerSerAsnLleTyrSerProLaserValSerProGlyAlaSer 269
OY 584 -----GCTGAATGTATTCCTGCG-----GGTTAAGAGGCTGATCTGCCGCAAT 539
DB 269 erAspGlyAla-SerAsnValProGlyLleGlyTyrSerGlyAlaLeuValAlaSerAsn 288
OY 538 -----TTGATATGACAGTTCTGTCTGCATCTGTGAAGACCAACTGCAGT 494
DB 289 AsnSerProTyrMetAlaAlaAlaAlaAlaAlaAlaHisAlaHisSerAsnTyrSer 308
OY 493 CATCAAGTGGTTTCCAGCTTCAAGTACTCTGTGGTGTGTTGAAGTCAATCAGCTGAGC 434
DB 309 AlaThrSerAsnAlaGln-----ValThrSerGly 318

```



```

Db 496 LeuThrThrSerLeuAlaGluLysGluGlnGlnThrAlaGlnLeuHisnLeuGlnThr 515
QY 353 TCTCTTACCACATATATACAAAGTTACTTGGAAATGAATTCAGAGAGAAATGAA 412
Db 516 GlnIleTyrgIleMet-----GluValGluLysGluGlnLysValGlu 529
QY 413 GTGTGGCAGGTGATCATATGAGACCAGAGCTGATGACTTCAAAACACCAAGACTTGG 472
Db 530 LeuValLysValGlnLeuGlnGlnAlaAlaGlnSerSerSerAlaGlnLysValLeu 549
QY 473 AAGTCGAACCACTTGGATGATGCA-----GTTGGCTTTTCAGAGATGGCAGCAGAA 526
Db 550 ArgAlaGlnIleGluGlnLeuGlnAlaLysLeuLysAlaValGlnGlnAlaLysAlaGlu 569
QY 527 GCTGATATCAACTGGCGGAGATAGGCTTATACCCGCGGAAATCAC----- 577
Db 570 AlaLeuAsnSerLeuLeuAlaGlnLysGlnHisLeuGlnAlaGlnLeuHisGlnLeuGly 589
QY 578 -----ATTCACTGGTGAAACTGCAAGTCGAGAGAGGTGCACAGCTC 619
Db 590 ValGluLysGlnGluLysLeuGlnMetValLysValGlnLeuGlnGlnAlaAlaGlnSer 609
QY 620 TCCCGGAAAGCAGAAACCAAGCTGGCAGAGCAGATAGAGAGCTCCGTCAGAAACA 679
Db 610 SerSerSerValGlnGlnAlaLeu---ArgAlaGlnIleGlnLysLeuGlnAlaLysLeu 628
QY 660 CAGCAGGAAGGGAGGAGCGC-----GCTGATGTCGAGCAG 715
Db 629 GlnGlnIleGlnGlnLysLysAsnAlaLeuAsnAlaSerLeuAlaLysGlnGln 648
QY 716 GAG---GGCTACCTGCTGAG 733
Db 649 GlnThrAlaGlnIleGlnGln 655

```

RESULT 7

069088 PRELIMINARY; PRT: 3164 AA.

```

AC 069088:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Virion protein.
OS human herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
CX NCBI_TaxID-10298;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-17;
RX MEDLINE-88274327; PubMed-2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McIad D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).
DR EMBL: D10879; BAA01682.1; -.
DR InterPro: IPR005210; Herpes_U136.
DR Pfam: PF03586; Herpes_U136; 1.
SQ SEQUENCE 3164 AA; 335817 MW; 2E3335F9525F8C71 CRC64;

```

Alignment Scores:

```

Pred. No.: 0.806 Length: 3164
Score: 112.50 Matches: 86
Percent Similarity: 33.44% Conservative: 21
Best Local Similarity: 26.88% Mismatches: 128
Query Match: 4.66% Indels: 85
Db: 12 Gaps: 17

```

US-09-939-293-1 (1-1358) x 069088 (1-3164)

QY 1252 CAGCTGTCTCTCAAGACAGTGGGGCAGAGTCA-----GAG 1217
 ::::::::::: ||::::::::::::

```

Db 2581 GluLeuAlaSerAspSerSerGlyProAlaProLeuAlaProAspProLeuSer 2600
QY 1216 AACACATCAGAAATACATACAAAGAAATCGTCAAAATGGAGTCCGCCCTCC 1157
Db 2601 ProThrAlaAspGlnSerValProThrSerGlnCysAlaProArg-ProProGlyProAl 2620
QY 1156 CACAACT-----GGCATCCCAACAGGAGCAAGACTACTAAATCAAT 1115
Db 2620 aValThrAlaArgGlnAlaAlaArgProGlyValProAlaGlnSerThrArgProAla----- 2638
QY 1114 TTTGACGACGTAAATTAAGACTGMAAAACAGTTTAAACACTTGCTGAACCTTAAGGGCATGAC 1055
Db 2639 -----ProValGlyProArgAspAspPheAr 2647
QY 1054 AAAA---AGGACTCTCTCTGACCCAGATAGGCAAAATGCTTTGGGTGATAGATAAA 998
Db 2647 gArgLeuProSerProGlnSerSerProGlyProAspAlaThrAlaProArgProPr 2667
QY 997 AAAATGGGTAAAGACAGCTGTACAGAGTGGGGTGAATGTTAAACGGGTGCAGTCCCA 938
Db 2667 cAlaSerSerAlaGln-----SerAlaAlaSerSerSe 2678
QY 937 AGGCTTAAGAACCAAGTCCAGCCAGCCAGCTGAGACCAGAGGACATCAGACTCACA 878
Db 2678 rGlySerAlaGlnAlaArgGlnAlaArgAlaArgSerLeuAlaArgAlaThrGlnAlaSe 2698
QY 877 AGGCTTCGCTGATTTGGCCAGGAGAGACTGCCGCTCTTCGCGTGACACAGACT 818
Db 2698 rAlaThrThrGlnGlnLysTrp-----ArgProAlaAlaLeu---ProAspThrVa 2713
QY 817 CATGCCAACCCCTGGGCGAGGGTG---GCATCTGCCCTGTTTCCACCTAGCTGGGAGAA 761
Db 2713 AlaProValThrAspPheAlaAlaArgProProAlaProProLysProProGlnProAlaPr 2733
QY 760 CAGGCGAGTGTCTCAGGCGCTC-----AATCTTCACCGCAGGTAGCGCTCTG 713
Db 2733 cHisAlaLeuValSerGly-ValProLeuProLeuGlyProGlnAlaAlaGlnAlaAlas 2753
QY 712 CTCGCACATCAGCCGCTCCTCCCTCTCTGTTGTTTTCAGCGAGCTCTTATCTG 653
Db 2753 eProAlaLeuProThrLeasProValProProAlaAlaThrGly-----T 2769
QY 652 TCGTCTGCGCAGCTTGATT-----CTGCTTCGGGAGAGCTGGCGAC 608
Db 2769 hrValLeuProGlyGlyLysAsnAlaArgProProLeuThrSerGly-----ProAlaP 2787
QY 607 CTCTTCACCTGCACTTTCACCCAGCTGAATGTGATTCCTGGCGTTATAGAGCGCTGATC 548
Db 2787 rThrProProAlaArgVal---ProValGlyGlyProGlnAlaArgGlyLeuThrArgPro---- 2804
QY 547 TCGCGCAGTTGATATGACAGCTTCGTCGCCATCTCTGAAAGACCACTGACGTCA---- 492
Db 2805 -----AlaValAlaSerLeuSerLysSerArgGlnSerLeuP 2817
QY 491 --TCCAAGTGGTTTCACGCTTCAATACCTCTGTTGTTT----- 453
Db 2817 rSerProThrAspProAlaAspProThrAlaProValLeuGlyLysAsnProAlaGlnP 2837
QY 452 --AAGTCATCTCAGCTGTGGCTCTATGATCAGCTGCACACACTTATCTCTCT 399
Db 2837 rThrSerSer-----SerProAlaGlyProSerProProPro 2849

```

RESULT 8

0946J0 PRELIMINARY; PRT: 332 AA.

```

AC 0946J0:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Putative retrotransposon protein.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

OC Ehnhartioidaeae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE:
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Galst'ner T., Riggs F., Hsiao J., Zisman V., Blunt S., Pal G.,
RA Vanaken S.E., Utteback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSUBB0031G04 genomic sequence.";
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC090714; AAK92683.1; -
SO SEQUENCE 332 AA; 35100 MW; A552B2BD31C7D4FC CRC64;

Alignment scores:	
Pred. No.:	0.596
Score:	111.50
Percent Similarity:	36.79%
Best local Similarity:	26.42%
Query Match:	4.62%
DB:	10
US-09-939-293-1 (1-1358) x Q94GJ0 (1-332)	
	Length: 332
	Matches: 51
	Conservative: 20
	Mismatches: 59
	Gaps: 63
	Indels: 6

OY	899	AGGAGGACCTCAGACGCTCACAAAGGCTTCGGCTGATTTGGCCAGGACGACTGGCG	84.0
Db	139	Argmeth1SerAspLeu1nHn1Ser1yAla1eu5nAsp1yLys1Leu1Val1Arg	158
OY	839	CTCTTCGGGTGCACAGACAGACTATGCGCAACCCCTGGGACAGGTGGACATCGCCCTGCTT	78.0
Db	159	Ile1leu1Ala1Leu1yAsp1Glu-----Met 166	
OY	779	TCCCACTGAGTGGGAGACAGGCGAGCTTGCTCAGGCCCTCAATCTCTACGACGAGTAG	72.0
Db	167	Ala1Arg1Leu1yLys1Glu1nAsn1Ala1nLeu1ySgl1yLeu1Pro1Ala1Pro1Arg1yVal1Arg	186
OY	720	-----	72.0
Db	187	Ile1Arg1Thr1Thr1Pro1Arg1yLys1Leu1Thr1Ala1Pro1Val1Arg1Ile1nLeu1Ala1Pro1Arg	206
OY	719	---CCCTCGCTCGCGACTCAGACCCGCGTCGCTCCGCTCTGCTGCTGTTTCTGCAGGAGCT	66.3
Db	207	Asn1Pro1Pro1Pro1Ala1Val1Pro1Ala1Ala1Pro1Ala1Pro1Ala1-----	22.9
OY	662	CTTCTATCTGTGCTTCTTGCCACAGCTTGGTTCTGCTTTCGGGAGAGCTGCTGCACTCTT	60.3
Db	222	-----Val1Pro1Ala-----Ala1Pro1Val 227	
OY	602	GCACCTGAGTTTCACACGAGTATGATTCCTCGGGGGTATATAGAGCCCTGATTCGCGC	54.3
Db	228	Pro1Pro1Ala1Val1Pro1Ala1Ala1Pro1-----Ser1Pro1Ala1Leu1Ala 240	
OY	542	CAGTTTGATATGACGACTTTCGCTGCCA-----TCTCTGAAGACCAACTGACATCTCC	48.9
Db	241	Leu1Val1Pro1Val1Pro1Val1Leu1Ala1Pro1Ala1Ser1Ala1Leu1Ser1Phe1nAla1Pro1Ala1Ser1Ala 260	
OY	488	AAGTGGTTTCCAGCTTCAAGTACTCTTGGTGTTTTGAAGTCAATCTCAGCTCTGGCTCTTA	42.9
Db	261	Ala1Arg1Gly1Pro1Ala1Ser1Gly1Asn1Gly1Trp1Leu1Ser1Ala1Thr1-Pro1Ser1Gly1Ser1Ar 280	
OY	428	TGATCACCCTGCACACTTCATCTTCCTCCCTCTGAA 394	
Db	280	gAsp1His-Ser1Ser1Ser1Ser1Ser1Arg1yIle1yThr1Glu 291	

RESULT 9	
061078	
ID 061078	PRELIMINARY; PRT; 520 AA.
AC 061078;	
DT 01-NOV-1996 (TREMBLrel. 01, Created)	
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)	
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)	
DE Wiscott-aldrich syndrome protein homolog.	

GN UA2471 OR N-3AP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mutelestomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxID=10090;

Alignment Scores:	
Pred. No.:	0.658
Score:	11.50
Percent Similarity:	42.86%
Best Local Similarity:	36.51%
Query Match:	4.62%
DB:	11
length:	530
Matches:	46
Conservative:	8
Mismatches:	51
Indels:	21
Gaps:	5

[illegible]

RESULT 10		
Q9ULE4		
ID	Q9ULE4	PRELIMINARY; PRT; 1068 AA.
AC	Q9ULE4;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)	
DE	KIAA1276 protein (Fragment).	
GN	KIAA1276.	
OS	Homo sapiens (Human).	

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K., Kikuno R., Hirosewa M., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:337-345(1999).
 DR EMBL; AB03102; BAA86590.1; -
 FT NON-TER
 SQ SEQUENCE 1068 AA; 121816 MW; 0BA2BCDC62CC2P95 CRC64;

Alignment Scores:
 Pred. No.: 1.26 Length: 1068
 Score: 109.00 Matches: 60
 Percent Similarity: 39.60% Conservative: 39
 Best Local Similarity: 24.00% Mismatches: 95
 Query Match: 4.46% Indels: 56
 DB: Gaps: 11

US-09-939-293-1 (1-1358) x Q9ULE4 (1-1068)

OY 254 TCTTGTGTAACAGATGACCTCTACCTTCTCTCT-----CAGACCACA 298
 Db |||||: : : |||||
 OY 299 TATGCTGTATGAAAGCT--ATTACTGAATATACAGGCTGTATACCTTAACCTCT 355
 Db ||| : : : ||| : : : ||| : : : |||||
 Db 610 LysAlaLysLeuGlnAlaGlnAlaSerGlnMetGlnAlaLeuGlnCysThrSer 629
 OY 356 CTTTACCGCAATATACAAAGTTACTTGGAAAATGAATTCAGAGAGAGAT----- 409
 Db |||||: : : ||| : : : ||| : : : |||
 Db 630 AsuTyArgGluAspLeuGlnAlaLeuLysGlnLeuSerAspLeuGlnArgGlnLysLeu 649
 OY 410 -----GAGGTGGCGAGTGATGATGAGCGCCAGCTGAGTGAATCTCA 454
 Db |||||: : : ||| : : : ||| : : : |||
 Db 650 GlnHisGlnLeuGlnCysThrThrGlnGlnAlaMetCysAlaGlnLeuGlnAla 669
 OY 455 AAACACCAAGAGTACTTGAAGCTGGAACACACTTGATGACTCACTGGCTTTGCAGAG 514
 Db |||||: : : ||| : : : ||| : : : |||
 Db 670 SerHisGlnArgAlaLeuArg----- 676
 OY 515 ATGGCAGCAGAGCTGCATATCAAACTGGCGCAGATCAGGCTCTATACCGCAGAGAT 574
 Db ||| : : : ||| : : : ||| : : : |||
 Db 677 MetLeuGlnLysAlaArgHisGln-----GluLeuLysAlaThrGln 690
 OY 575 CACATTCACTGTGTGAAGCTGGAGGAGAGTGCACACAGCTCCCGGAAGACAGAA 634
 Db |||||: : : ||| : : : ||| : : : |||
 Db 691 Glu-----ArgLeuLysLysGlnSerSerHisLeuGlnHisGln 706
 OY 635 ACCAAGCTGGCAGAACACAGATAT-----GAAGAGCTCCGTCAGAAAACACAGAGGAA 688
 Db ||| : : : ||| : : : ||| : : : |||
 Db 707 ThrHisArgLeuGlnLeuGlnAlaLeuGlnLysAlaArgGlnLysLeuGlnGln 726
 OY 689 GGGGAGAGCGGCTGAGTCGAGCAGCAGAGGCTTAC-----CTGCGTGAAGATTGA 739
 Db |||||: : : ||| : : : ||| : : : |||
 Db 727 --ArgGluArgMetGlnAlaGlnGlnAlaLeuLeuLeuGlnSerLeuArgGlnGlnLeu 745
 OY 740 GGGCGTGCAGACACTGCTCCCTGCTCCCACTGAGTGGGAAACAGAGGGGAGATGGCACC 799
 Db : : : ||| : : : ||| : : : |||
 Db 746 SerGlnGlnGlnAlaAlaLysSerGlnHisGlnLysAspLeuGlnAlaLeuGlnAlaGln 765
 OY 800 CTGCCAGGGGTGGCATGACTGTCTGTGCAACCGAAGAGCGCAGATGCC----- 850
 Db ||| : : : ||| : : : ||| : : : |||
 Db 766 LeuArgAlaLeuGlnArgGlnGlnAlaSerSerGlnCysProGlnLysSerLysAspHis 785
 OY 851 -----TGCCCTGGCCAAATCAGCGCAGACGCCCTTTGTAGCT 886
 Db |||||: : : ||| : : : ||| : : : |||
 Db 786 IleIleAlaThrGlnGlnLysArgGlnGlnLysArgGlnAlaGlnSerProGlnAlaAla 805

OY 887 GTGAGTGCCTCC-----TGTGCTCTC 907
 Db ||| : : : ||| : : : ||| : : : |||
 Db 806 GlyGlnGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 815

RESULT 11
 ID Q9SUC7 PRELIMINARY; PRT; 1396 AA.
 OX Q9SUC7
 AC Q9SUC7
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Hypothetical 154.2 kDa protein.
 GN T13K14.10 OR AT4G20850.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Lemcke K., Schueller C.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL080282; CAB45880.1; -
 DR EMBL; AL161553; CAB79083.1; -
 DR HSP; P00782; 2Sbr.
 DR MEROPS; S08.090; -
 DR InterPro; IPR00209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 3.
 DR PRINTS; PR00723; SUPRTILASIN.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1396 AA; 154174 MW; 113209E42B9DEAE3 CRC64;

Alignment Scores:
 Pred. No.: 1.78 Length: 1396
 Score: 107.50 Matches: 48
 Percent Similarity: 36.15% Conservative: 29
 Best Local Similarity: 22.54% Mismatches: 71
 Query Match: 4.39% Indels: 65
 DB: Gaps: 6

US-09-939-293-1 (1-1358) x Q9SUC7 (1-1396)

OY 203 AAATCAGAGCCCTGATCCCT----- 223
 Db ||| : : : ||| : : : ||| : : : |||
 Db 1128 LysAspAsnProHisArgLeuValLysLeuAspAlaProGlnLysAspLysAlaAla 1147
 OY 224 -----AGTAGTGAAGCATGATGAGAGAGAGCTGCTTGGTACAGATAGC 271
 Db |||||: : : ||| : : : ||| : : : |||
 Db 1148 SerAlaProThrCysSerLysSerValSerGlnArgLeuGlnGlnLysValAlaArgAspThr 1167
 OY 272 ACCTCACTTCTCTCTGACACACATATGCGTGTATGAACTTACTGAAATPACT 331
 Db |||||: : : ||| : : : ||| : : : |||
 Db 1168 LysIleLysPheLeuLysLeuLysGlnGlnThrGlnGlnLysArgSerGlnThrArg 1187
 OY 332 AAGCGTGTATACCTTAATCTTCTTACCGCAATATACAGTTACTTGGGAAATG 391
 Db ||| : : : ||| : : : ||| : : : |||
 Db 1188 LysLeuCysThrCysLeuLysSerGlnLysProAspThrThrProLeuLeuAlaLysIle 1207
 OY 392 -----AATTCAGAGGAGGAAGATGAGAGTGGCAGGTGATCATAGCA 433
 Db |||||: : : ||| : : : ||| : : : |||
 Db 1208 LeuGlnGlnLeuLeuSerArgSerAspAlaGlnLysAspLysIle----- 1221

```

QY 434 GCCAGAGCTGAGTACTTCAAAACACCAAGAGTACTTGAAGCTGGAAACCACTTGGATG 493
Db 1222 -----SerHisHisGluGluIleIleGluAlaAlaAsnGluValVal 1235
QY 494 ACTGACATGGTCTTTCAGAGATGGCAGACAGAGCTGCATATCAAACTGCCAGATCAG 553
Db 1236 ArgSerValAspValAspGluLeu-----1243
QY 554 GCCTGTATACCGCAGCAATCATTCAGCTGTGAATACGACAGAGAGGTCGAC 613
Db 1244 -----AlaArgPheLeuLeuAspLysThrGluProGluAspAspGluIleGlu 1259
QY 614 CAGCTCTCCGCGAAGACAGAAACCAAG-----640
Db 1260 LysLeuLysLysMetGluValThrArgAspGluLeuAlaAspAlaLeuTrpGlnLys 1279
QY 641 ---CTGGCAGAAACACAGATAGAGAGCTCCGTCAGAAAACACAGAGAGAGGAGAG 697
Db 1280 GlyLeuAlaMetAlaArgIleGluAsnLeuLysGlyGlyGluGlyGluGlyGlu 1299
QY 698 CGGCTGAGTGGAGGAGGAGGCTACCTCGCTGAGGAT 736
Db 1300 -----GluSerSerGlnLysAspLysPheGluGluAsn 1310

RESULT 12
Q924W1 PRELIMINARY; PRT; 356 AA.
ID 0924W1
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALEX protein.
GN ALEX.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95089824; PubMed=7997272;
RX Kehlenbach R.H., Matthey J., Huttner W.B.;
RX "Xlas is a new type of G protein.";
RL Nature 372:804-809(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Klemke M., Kehlenbach R.H., Huttner W.B.;
RT "Two overlapping reading frames in a single exon encode interacting
RT proteins - a novel way of gene usage.";
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X84047; CAC39212.1;
SQ SEQUENCE 356 AA; 37970 MW; 9849ABD0AE524A3D CRC64;

Alignment Scores:
Pred. No.: 1.46 Length: 356
Score: 107.00 Matches: 46
Percent Similarity: 38.73% Conservative: 9
Best Local Similarity: 32.39% Mismatches: 45
Query Match: 4.43% Indels: 42
DB: Gaps: 8

US-09-939-293-1 (1-1358) x Q924W1 (1-356)
QY 946 CAGTCCCAAGGGCTAAGACAGCTCCAGCAAGCTGAGACCAAGAGGACACTCAG 887
Db 72 GlnSerProThrProLys-----ArgSerProGln-ProAlaGlnProLeuPr 87
QY 886 AGCTCAACAAGCGCTGCTGATGGCCAGGCGAGGAGCCTG-----844
Db 87 OAArgArgSerLeuProGlyGlnProSerProLeuArgSerProLeuProG1 107
QY 843 -----CCGCCTTCTCGGTGTCAGACAGACGTAT 815

```

```

Db 107 yLeuSerLeuLeuProGluProIleGlnProGlyLeu-SerLeuGluProGlnArgC 127
QY 814 GCCAACCCCTGGGAGGGTGGCATCTGCCCTGCTTCCCACTG-----AGTG 767
Db 127 yGlnProLeuLeuGly-----GlnProProLeuGluGlnProMetGlnValLeuTrpSerG 146
QY 766 GGGAGACAGGCGCATGCTCTCAGAGCCCTCAATCCTCAGCAGAGTAGGCTCTGCTCGGA 707
Db 146 yGluProGlyHis---SerArgLeuLeuGlnPro---LeuGlyHisProSerLeuProA 164
QY 706 CTCAGCCCGCTCCCTCCCTCCCTGCTTCTTGAGAGAGCTCTTATCTGCTTC 647
Db 164 IagInGlnLeuProProGluGlnPro-----LeuL 174
QY 646 TGCCAGCTTGGTCTCTGCTTCCGAGAGAGCTGTCACCTTCCACCTGACATTTTCAC 587
Db 174 euProIaGlnSerLeu-----ProIaGlyGlnProLeuProProGlnIaGlyP 191
QY 586 CA 585
Db 191 ro 191

RESULT 13
Q99M73 PRELIMINARY; PRT; 603 AA.
ID 099M73
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Type II 65kd keratin.
GN KRT2-16.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C3H/HEN;
RA Polier C.; Yoshiki A.; Fujiwara K.; Guenet J.-L.; Kusakabe M.;
RT "Fine mapping of Hague (Hag), a mouse hair mutation with 2
RT epialleles.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY028607; AKR26233.1;
DR MGD; MGI:96700; Krt2-16.
DR InterPro; IPR001654; IF.
DR InterPro; IPR003054; keratin_II.
DR Pfam; PF00038; filament_1.
DR PRINTS; PRO1276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; UNKNOWN_1.
SQ SEQUENCE 603 AA; 64987 MW; AFC01BD8D0E780B9 CRC64;

Alignment Scores:
Pred. No.: 1.64 Length: 603
Score: 107.00 Matches: 42
Percent Similarity: 47.70% Conservative: 41
Best Local Similarity: 24.14% Mismatches: 59
Query Match: 4.37% Indels: 32
DB: Gaps: 7

US-09-939-293-1 (1-1358) x Q99M73 (1-603)
QY 251 GTGCTTTGGTAACAGATAGCAGCTTACTTCTCTCTCAGACCACTAT---GCGTTG 307
Db 276 ValAlaLeuLysLysAspValAspAlaAlaPheLeuAsnLysSerAspLeuAlaAsn 295
QY 308 ATTGAAGCTAT-----ACTGAATATACTAAGCTGTTTAT-----ACC 346
Db 296 ValAspThrLeuIleGlnGluThrGluPheLeuLysAlaLeuTrpHisGluGluIleGlu 315
QY 347 TTAACCTCTCTTACCGCAGCATATACAGAGTTACTTGGGAAATGAATTCAGAGAGAA 406
Db 316 MetLeuGlnSerHisIleSerGluThrSerValIleValLysMetAspAsnSerArgAsp 335

```

```

QY 407 GATGAAGTGTGGAGGTGATCATAGAGCCAGAGCTGATGACTTCAAAACACAGAG 466
    ::::: |||||::: ::::: |||||:::
Db 336 LeuAnleuaspGlyIleIle-----AlaGluValIysAlaGlnTyrGluGlu 351
QY 467 TACTTGAAGTGGAAACCACTTGGATGACTGCTGTTGTTTTCAGAGATGACGAGAA 526
    ::::: |||||:::
Db 352 ValAlaIArg-----ArgSerArgAlaAspValGlu 361
QY 527 GCGCATATCAACAGTGGCCGAGATCAGGCGCTTATACCCGCCAGATTCATTCAGCTG 586
    ::::: |||||:::
Db 362 SerTrpTyrGlnThrIlystYrGluGluMetArgValThrAlaGlnIlystYsAspAsn 381
QY 587 GTGAACCTGAGGTGGAGAGAGTGCACACAGCTCTCCCGG-----AAAGCA 631
    ::::: |||||:::
Db 382 LeuArgSerThrArgAspGluIleAsnGluLeuThrArgLeuIleGlnArgLeuIysAla 401
QY 632 GAACCAAGCTGGCAGACAGACAGATAGAAAGCTCCGTCAGAAACACACAGAGAGGG 691
    ::::: |||||:::
Db 402 GluIleGluThrIlystYrGluGlnIlystYsAlaIysLeuGluValAlaValAlaGlu 419
QY 692 GAGAGGCGGTGAGTGGCAGAGAGAGGCGCTTACTGCTGAG 733
    ::::: |||||:::
Db 420 -----AlaGluGlnGlnIlyGluAlaIleAsnAsp 430

RESULT 14
092213 PRELIMINARY; PRT; 738 AA.
AC 092213;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Hypothetical 80.3 kDa protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RA Wang Y.Z., Kehlenbach R.H., Huttner W.B.;
RT "The XL-domain of rat xias is encoded by a single exon.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF093569; AAD03033.1;
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENSX.
KW Hypothetical protein.
SQ SEQUENCE 738 AA; 80340 MW; 51EA2B3A7D9D018A CRC64;

Alignment Scores:
Pred. No.: 1 71 Length: 738
Score: 107.00 Matches: 46
Percent Similarity: 38.73% Conservative: 9
Best Local Similarity: 32.39% Mismatches: 45
Query Match: 4.43% Indels: 42
DB: 11 Gaps: 8

US-09-939-293-1 (1-1358) x 092213 (1-738)
QY 946 CAGTGCACAGGCGCTAAGAACAGGTCCAGAGCCTGAGACACAGAGGAGCATCTAC 887
    ||| ||| |||
Db 454 GlnSerProThrProIly-----ArgSerProGln-ProArgGlnProLeuPr 469
QY 886 AGCTCACAAAGCGCTCTCGGCTGATTGGCCAGGCGCAGGACTG-----844
    ::::: |||||:::
Db 469 oArgArgArgSerLeuProGlnIlyProIlyProIlyProIlyProIlyProIlyProI 489
QY 843 -----CCGCTCTTCTCGGTCACACAGAGCATCTAC 815
    ||| ||| |||
Db 489 yLeuSerLeuLeuProGluProIleGlnProProGluIlyLeu-SerLeuGlnProGlnIly 509
QY 814 GCCAACCTGGGAGGCGTGCATCTGCGCTTTCGCCACTG-----AGTG 767
    ||||| ||| |||||
Db 509 yGlnProLeuLeuGly-----GlnProProLeuGlnIlyProMetGlnValLeuTrpSerG 528

```

```

QY 766 GGGAGACAGGCGAGTGTGTCTACAGCCCTCAATCTCAGCAGGTAGGCGCTGCTCGA 707
    ||||| ||| ::::: ||| ||| ||| |||
Db 528 IyGluProGlnIlyS-----SerArgLeuLeuGlnPro-----LeuGlnIlySProSerLeuPro 546
QY 706 CTCAGCCCGCTCTCCCTCTCTCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 647
    ||| ||||| |||
Db 546 IacGlnIlyLeuProProGlnIlyPro-----LeuL 556
QY 646 TGGCAGTTGTTTCTCTCTTCCGGGAGAGCTGTGACCTCTTCCACTTCGACTTTCAC 587
    ||||| ||| ||||| ||||| |||||
Db 556 euProIacGlnSerLeu-----ProAlaGlnIlyGlnProLeuProProGlnIlyPro 573
QY 586 CA 585
    ||
Db 573 ro 573

RESULT 15
09ROT4 PRELIMINARY; PRT; 739 AA.
ID 09ROT4;
AC 09ROT4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 84.0 kDa protein (Fragment).
GN pSPC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EI34;
RA MEDLINE=20038319; PubMed=10569772;
RA Brooks-Walter A., Billes D.E., Hollingshead S.K.;
RT "The pSPC gene of streptococcus pneumoniae encodes a polymorphic
RT protein, pSPC, which elicits cross-reactive antibodies to pSPa and
RT provides immunity to pneumococcal bacteremia.";
RL Infect. Immun. 67:6533-6542(1999).
DR EMBL: AF068647; AAF13457.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF01473; CW_binding.1;
DR PRINTS: PRO1574; TUBBYPROTEIN.
DR PRINTS: PRO1217; PRICHEXTENSX.
DR TIGRfams: TIGR01106; YSIRK_signal; 1.
KW Hypothetical protein.
KW NON_TER
FT NON_TER 739
SQ SEQUENCE 739 AA; 83961 MW; 7EE2FE676ABF989 CRC64;

Alignment Scores:
Pred. No.: 1 71 Length: 739
Score: 107.00 Matches: 51
Percent Similarity: 40.09% Conservative: 40
Best Local Similarity: 22.47% Mismatches: 70
Query Match: 4.37% Indels: 66
DB: 2 Gaps: 8

US-09-939-293-1 (1-1358) x 09ROT4 (1-739)
QY 203 AAATCAGAGCGCTCATTCCTTAGTAGAGCATTTGATGAGAGGAGGAGTGTCTTGTA 262
    ||||| ||| ||||| |||
Db 291 LysSerSerAspSerSerValIlyGluGlnIlyThrLeuProSerProSerLeuAsnMeLa 310
QY 263 ACAGATAGCAGCTCTACCTTCTCTCTCAGACACCATATGCGTTGATGAAGCTATTACT 322
    ::::: |||||:::
Db 311 AsnGlu-----SerGlnThrGlnIlyIlyArg-----LysAspValAsp 322
QY 323 GAATATACTAAGCGCTTTTACTTACTTCTTTCACGACAA-----367
    ||||| ||| ::::: ||| |||||
Db 323 GluIlyIlyLysIlyMetLeuSerGluIleGlnLeuAspArgIlySHIstHlGlnAsn 342
QY 368 -----TATACAGTTTACTT 382

```

```
Db 343 ValAsnLeuAsnIleLysLeuSerAlaIleLysThrLysTyrLeuTyrGluLeuSerVal 362
OY 383 GCGAAATGATTTCAGAGAGAGAAATGATGAGGTGATCATAGAGAGAGCT 442
Db 363 LeuLysGlnSerLysLysGlu-----GluLeuThrSerLysThrLysAla 378
OY 443 GAGATGACTTCAAAACACCAAGAGTACTGAAAGCTGGAACCACTTGATGACTGCAGTT 502
Db 379 GluLeuThrAlaIaPheGluGlnPheLysLysAspThrLeuLysProGluLysLysVal 398
OY 503 GGTCTTTCAGAGATGCGAGAGAGCTGCATATCAAACTGCGCAGATCAGGCTCTATA 562
Db 399 AlaGluAlaGluLysLysValGluGluAlaLysLysAlaLysAspGlnLysGluGlu 418
OY 563 ACCGCCAGAAATCAC-----ATTACAGCTGGTGAACCTGCAGGTG----- 577
Db 419 AspArgArgAsnTyrProThrAsnThrTyrLysThrLeuGluLeuGluIleAlaGluSer 438
OY 578 -----ATTACAGCTGGTGAACCTGCAGGTG----- 601
Db 439 AspValLysValLysLysAlaGluLeuGluValLysGluGluAlaAsnGluSerArg 458
OY 602 -----GAAGAGGTGCGACAGCTCTCCGGAAAGCAGAAACCAAGCTGCGCAGAAGCA--- 652
Db 459 AsnGluGluLysLysLysGlnAlaLysGluLysValGluSerLysLysAlaGluAlaThr 478
OY 653 -----CAGATAGAAAGACTCCGTACAGAAACACAGCAGAGAGGAGAGCGGCT 703
Db 479 ArgLeuGluLysLysLysThrAspArgLysLysAlaGluGluGluAlaLysArgLysAla 498
OY 704 ---GAGTCGAGCAGCAGAGGCC 721
Db 499 GluGluSerGluLysLysAla 505
```

Search completed: February 20, 2003, 15:01.17
Job time : 107 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	1201	49.1	239	21	AA026210	Human caspase acti
2	1201	49.1	239	21	AA078447	Inhibitor of apopt
3	1138	46.5	227	21	AA054139	Human pancreatic c
4	879	35.9	186	22	AA092922	Human protein sequ
5	190	7.8	40	23	AA078430	Inhibitor of apopt
6	186	7.6	39	23	AA078436	Inhibitor of apopt
7	164	6.7	35	23	AA078439	Inhibitor of apopt
8	141	5.8	30	23	AA078435	Inhibitor of apopt
9	125.5	5.2	502	22	AA053315	Human WASP protein
10	125.5	5.2	502	22	AA067333	Amino acid sequenc
11	125.5	5.2	502	22	AA067333	Human Wiscott-Aldr
12	119.5	5.0	317	22	AA040791	Human polypeptide
13	119.5	5.0	317	22	AA040792	Human polypeptide
14	116	4.8	276	22	AA037672	Novel human diagno
15	112	4.6	1485	21	AA059166	A mature human N-a
16	111.5	4.6	520	22	AA052318	Murine WASP protei
17	111.5	4.6	520	22	AA067330	Amino acid sequenc
18	111	4.5	401	22	AA067330	Novel signal trans
19	110	4.5	589	22	AA017136	Drosophila melanog
20	108	4.4	525	23	AA058670	Human DITP polype
21	107.5	4.4	1396	23	AA060120	Heridically activ
22	107	4.4	589	20	AA043392	PSPC alpha-helix c
23	107	4.4	929	18	AA014563	Streptococcus pneu
24	107	4.4	929	20	AA043384	S. pneumoniae Papc
25	107	4.4	1509	23	AA079051	Human zneu2 polype
26	106.5	4.4	344	22	AA012044	Human protein SEO
27	106.5	4.4	370	22	AA080035	Human retinal pigm
28	106.5	4.4	370	22	AA049336	Murine WASP protei
29	106	4.4	520	22	AA048796	Drosophila melanog
30	106	4.4	591	19	AA059825	Drosophila melanog
31	104.5	4.3	670	22	AA059825	Human protein sequ
32	104	4.3	663	22	AA059825	Human protein sequ
33	104	4.3	749	22	AA062915	Human protein sequ
34	103.5	4.2	317	22	AA087430	Drosophila melanog
35	101.5	4.1	1197	22	AA040254	Human gene 11 enco
36	101.5	4.1	1297	22	AA095107	Human polypeptide
37	101	4.1	441	22	AA095107	Human protein sequ
38	101	4.1	1072	23	AA095107	Human protein sequ
39	101	4.1	1072	23	AA095107	Human protein sequ
40	100.5	4.1	1084	23	AA095107	Human protein sequ
41	100	4.1	337	22	AA095107	Human protein sequ
42	100	4.1	518	22	AA095107	Human protein sequ
43	100	4.1	551	21	AA042049	Human ORFX ORP1813
44	100	4.1	564	19	AA062654	C3 binding protein
45	100	4.1	581	20	AA043394	S. pneumoniae SpnA

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Query Match: 4.588 Indels: 42
DB: 1 Gaps: 8

US-09-939-293-1 (1-1358) x S02041 (1-3660)

```

OY 5 TCCGCCGCTGCACAAATGCGCGCTGTGAAGAGTGTGCTGCGGACGCTACTTCATTC 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1163 SerLeuAaTgLyLeuAaPLeuSerLeuMetHisGluTrpIleHisGluGluGluTrp 1192
OY 65 TTCAGGTACAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1183 LeuGlu-----ArgAspPheGluTrpIleHisGluGluGluGluGluGlu 1192
OY 125 TCAGATTGATAGACATGACAAACTGTGAGATTGCGTTGGAGTAAACCTGTGT 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1193 GluGluLeuGluGlu----- 1197
OY 185 GCGGTCTCTATTGACAGAAATCAGAGCTCTATCCCTAGTACTGAGCATTTGATGAG 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1198 AlaValGluGluLeuLeuAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAla 1214
OY 245 AGAGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1215 ---LysValLysLeuIleHisPheSerValAsnAsnPheIleAlaLysAlaProProAla 1233
OY 305 TTGATTGAGCTATTACTGAATATATCTAAGCTCTTATCTTACTTCTTACTTCTTAC 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1234 AlaAsnGluAlaLeuLysLysGluLeuAspValLeu-----IleHisSerTrpGlnArg 1251
OY 365 CAATGACAAATTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1232 LeuCySerLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPLeu 1267
OY 425 ATCATAGAGCAGCAGCTGAGATGATCTTCAAAACACAGAGTACTGATGAGTGAACC 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1268 -----AlaCySerTrpHisGluLeuLeuSer-----TyrLeuAspAlaGluAsn 1281
OY 485 ACTGGATGATGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1282 LysTrpLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPLeu 1300
OY 545 GAGATGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1301 AlaGluGluIleSerGluSerLeuAspSerLeuGluTrpLeuMetArgHisProGluAsp 1320
OY 605 GAGGTGACACAGCTCTCCCGGAAACAGAAACCAAGCTGCGCAGAGACACATGAGAG 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1321 AsnArgAsnGlnIleArgGluLeuAlaGluIleHisPheGluGluIleLeuAspGlu 1340
OY 665 CTCGCTGAGAAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1341 LeuLeuAsnGluLysLeuGluLysPheAsnTrpArgTrpGluGluGluGluGluAla 1360
OY 725 CTCGCTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1361 ValArgArgGlnLysSerLeuGluGluSerIleGlnSerAlaGluGluTrpAspLys 1379

```

RESULT 5

178557 N-methyl-D-aspartate receptor chain NMDAR2D-2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Aug-1996 #sequence-revision 02-Aug-1996 #text-change 21-Jan-2000
C:Accession: I78557; I58158; D45219
R:Monyer, H.; Burnashev, N.; Laurie, D.J.; Sakmann, B.; Seeburg, P.H.
A:Title: Developmental and regional expression in the rat brain and functional properties of the N-methyl-D-aspartate receptor subunit NR2B
A:Reference number: I58158; MIMD:94206533; PMID:7512349
A:Accession: I78557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1323 <RES>
A:Cross-references: GB:I31612; NID:9469068; PIDN:AAC37647.1; PID:9469068
A:Accession: I58158

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-66, 'V', 68-1323 <RES>
A:Cross-references: GB:I31611; NID:9469066; PIDN:AAC37646.1; PID:9469067
R:Shi, T.; Moriyoshi, K.; Suda, H.; Sakurada, K.; Kadohira, M.; Yokoi, M.; Akai, J.
A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor subunit NR2B
A:Reference number: A45219; MIMD:94155102; PMID:8428958
A:Accession: D45219
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1265-1323 <ISH>
A:Experimental source: brain
A:Note: sequence extracted from NCBI database (NCBI:124265)
A:Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology F:451-879/Domain: glutamate receptor homology <GRH>

Alignment Scores:

Pred. No.	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
1	111.00	1323	56	8	73	67	9
2	31.37%	27.45%	4.60%	2	6	6	2

US-09-939-293-1 (1-1358) x I78557 (1-1323)

```

OY 915 GCAAGCTGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1004 AlaLeuValArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1023
OY 855 GGGCAGAGCAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1024 SerProProAlaProProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1043
OY 795 GCA-----TCGCCCTGCTTCCCACTGAGTGC----- 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1044 AlaPheGluAspGluSerProProAlaProSerArgTrpProArgSerAspProGluSer 1063
OY 765 -----GGAGACAGGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1064 GlnProLeuLeuGlyGlyGlyAla-----GlyGlyProSerAlaGlyAlaProThrAla 1081
OY 717 TCCGCTCCGACCTGAGCGCGCTCTCC-----CTTCCTCTGTGT 676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1081 AspProProAlaArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1101
OY 675 TTCGTAC-----GGAGCTCTTCTATCTGCTCTCTCCAGC 640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1101 oSerAspSerGluAspSerGluSerLeuGlyAlaSerLeuGlyGlyLeuGluProTrp 1121
OY 639 TTGGTTCTG-----CTTCCGGAGAGCTGTGACCT----- 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1121 PTrpPheAlaAspPheProTrpProTrpAlaGluAlaGluGlyProProGluAlaGly 1141
OY 605 -----CTTCACCTGC 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1141 TrpSerValAspLysLeuGlyTrpArgAlaGlySerTrpAspTrpLysLeuProTrp 1161
OY 594 AGTTTACCAAGCT-----GAATGTGATTCCTGGCGTTATAGAGCTGATCT 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1161 gGlyGlyProAlaTrpHisCysArgHisCysAlaSerLeuGluLeuLeuProProTrp 1181
OY 546 GCGCCAGTTTGATATGACAGTTTGTGTCATCTTGAAGACCAACTGACATATCAA 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1181 gHis-----LeuSerCysSerHisAspGlyLeuAspGlyGlyTrp 1194
OY 486 GTGGTTCCA 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1194 pTrpAlaPro 1197

```

RESULT 6

C45219

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 20, 2003, 14:47:03 ; Search time 73.5 Seconds

(without alignments)
4923,928 Million cell updates/sec

Title: US-09-939-293-1

Sequence: 1 ggcgtccgcgcgcgcgcacaa.....ggacttaacacagaaaaaa 1358

Scoring table:

BIOSM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame_plus_n2p_model -DEV-rlp
-O/-cgn2.1/USPNC_SPOOL/US09939293/runat_20022003.111509.14822/app_query.fasta.1.1543
-DB-A_Geneseq_101002 -OFMT-fasta -SUFFIX-frag -MIMMATCH-0.1 -LOOPCL-0
-LIST-45 -DOCALLIGN-200 -START-1 -END-1 -MATRIX-biosm62 -TRANS-human40.cdl
-MODE-LOCAL -OUTFMT-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15
-USER-US09939293.ecgn.1.1.47.etrnat.20022003.111509.14822 -NCPU-6 -ICPU-3
-NO_XLIPYX -NO_MMAP -LARGEQUERY -NEG_SCORES-0 -WAIT -LONGLOG -DEV-TIMEOUT-120
-WARN_TIMEOUT-10 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -Fgapop-6 -Fgapext-7
-YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database:

A_Geneseq_101002:*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1201	49.1	239	21	AA26210	Human caspase acti
2	1201	49.1	239	21	AA26210	Inhibitor of apopt
3	1138	46.5	227	21	AA26210	Human pancreatic c
4	879	35.9	186	22	AA26210	Human protein sequ
5	190	7.8	40	23	AA26210	Inhibitor of apopt
6	186	7.6	39	23	AA26210	Inhibitor of apopt
7	164	6.7	35	23	AA26210	Inhibitor of apopt
8	141	5.8	30	23	AA26210	Inhibitor of apopt
9	125.5	5.2	502	22	AA26210	Human WASP protei
10	125.5	5.2	502	22	AA26210	Amino acid sequenc
11	125.5	5.2	502	22	AA26210	Human WASP protei
12	119.5	5.0	317	22	AA26210	Human WASP protei
13	119.5	5.0	317	22	AA26210	Human WASP protei
14	116	4.8	276	22	AA26210	Human WASP protei
15	112	4.6	1485	21	AA26210	Human WASP protei
16	111.5	4.6	520	22	AA26210	Human WASP protei
17	111.5	4.6	520	22	AA26210	Human WASP protei
18	111	4.5	520	22	AA26210	Human WASP protei
19	110	4.5	589	22	AA26210	Human WASP protei
20	108	4.4	525	23	AA26210	Human WASP protei
21	107.5	4.4	1396	23	AA26210	Human WASP protei
22	107	4.4	589	20	AA26210	Human WASP protei
23	107	4.4	929	18	AA26210	Human WASP protei
24	107	4.4	929	20	AA26210	Human WASP protei
25	107	4.4	1509	23	AA26210	Human WASP protei
26	106.5	4.4	344	22	AA26210	Human WASP protei
27	106.5	4.4	370	22	AA26210	Human WASP protei
28	106.5	4.4	370	22	AA26210	Human WASP protei
29	106	4.3	520	22	AA26210	Human WASP protei
30	106	4.3	591	19	AA26210	Human WASP protei
31	104.5	4.3	670	22	AA26210	Human WASP protei
32	104	4.3	663	22	AA26210	Human WASP protei
33	104	4.3	749	22	AA26210	Human WASP protei
34	103.5	4.2	317	22	AA26210	Human WASP protei
35	101.5	4.1	1197	22	AA26210	Human WASP protei
36	101.5	4.1	1297	22	AA26210	Human WASP protei
37	101	4.1	441	22	AA26210	Human WASP protei
38	101	4.1	1072	23	AA26210	Human WASP protei
39	101	4.1	1719	23	AA26210	Human WASP protei
40	100.5	4.1	1084	23	AA26210	Human WASP protei
41	100	4.1	337	22	AA26210	Human WASP protei
42	100	4.1	518	22	AA26210	Human WASP protei
43	100	4.1	551	21	AA26210	Human WASP protei
44	100	4.1	564	19	AA26210	Human WASP protei
45	100	4.1	581	20	AA26210	Human WASP protei

ALIGNMENTS

RESULT 1	AA26210	standard; Protein; 239 AA.
AA26210	standard; Protein; 239 AA.	
XX	AA26210:	
XX	23-FEB-2001 (first entry)	
DT	Human caspase activator Smac.	
DE	Human caspase activator Smac.	
XX	Human caspase activator Smac.	
XX	Human caspase activator Smac.	
KW	Human: caspase activator; Smac; apoptosis; cancer; autoimmune disease; neurodegenerative disease; mitochondria.	
KM	Human: caspase activator; Smac; apoptosis; cancer; autoimmune disease; neurodegenerative disease; mitochondria.	
XX	Homo sapiens.	
OS	Homo sapiens.	
PN	US6110691-A.	
XX	US6110691-A.	
PD	29-AUG-2000.	
XX	29-AUG-2000.	

THIS PAGE BLANK (USPTO)

PF 06-JAN-2000; 2000US-0479309.
 XX
 PR 06-JAN-2000; 2000US-0479309.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Wang X, Du C;
 XX
 DR MPI: 2000-586350/55.
 XX N-PSDB; AAA94860.
 XX
 PT Novel caspase regulatory polypeptides useful for screening binding
 PT agents specific for the polypeptides which are useful for diagnosis and
 PT also for treating apoptosis associated diseases
 XX
 PS Claim 1; column 23-24; 16pp; English.
 XX
 CC The present sequence is the human Smac protein. Its coding sequence
 CC was isolated by purifying the protein and searching a HeLa cell cDNA
 CC library for sequences which bound to probes based upon its Smac is a
 CC mitochondrial protein which is released into the cytosol during
 CC apoptosis, and acts as a caspase-3 activator. The protein and its coding
 CC sequence can be used to modulate the expression and function of caspases
 CC and their activators, and also can be used as drug targets and regulators
 CC to promote or inhibit apoptosis in the treatment of cancer and autoimmune
 CC and neurodegenerative diseases.
 XX
 SO Sequence 239 AA;
 XX
 Alignment Scores:
 Pred. No.: 3.17e-105 Length: 239
 Score: 1201.00 Matches: 239
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 49.10% Indels: 0
 DB: 21 Gaps: 0
 US-09-939-293-1 (1-1358) x AAB26210 (1-239)
 QY 20 ATGGCGGCTGTGAAGAGTGGCTGTGGCGAGCGTACTTATCTTTCAGAGTACAGACAG 79
 DB 1 MetAlaAlaLeuLysSerTrpLeuSerArgSerValThrSerpheheargtryrarglin 20
 QY 80 TGTGTGTGTCCTGCTGTGGTGTACTTTAAGACGGGTGTTTCAGAAATGATGAAGA 139
 DB 21 CysLeuGlyValProValAlaAlaasnPhelLysLysArgCysPheSerGluLeuIleArg 40
 QY 140 CCATGGCACAATACTGTGACGATTGGCTTGGAGTAACCTGTGTGGCTGCTATTGCA 199
 DB 41 ProTrpHisLysThrValThrIleGlyPheGlyValThrLeuGlyAlaValProIleAla 60
 QY 200 CAGAAATGACAGCCCTCATTCCTTACTAGTACGATGATGATGAGAGACAGTCTTTG 259
 DB 61 GlnLysSerGluPheHisSerLeuSerSerGluAlaLeuMetArgLysAlaValSerLeu 80
 QY 260 GTAACAGATAGACCT 319
 DB 81 ValThrAspSerThrSerThrPheLeuSerGlnThrThrTyrAlaLeuIleGluAlaIle 100
 QY 320 ACTGATATCTACTAGGCTGTTTATACCTTAATCTCTTCTTACCGACATATCAAGTTTA 379
 DB 101 ThrGluTyrThrLysAlaValThrThrLeuThrSerLeuTyrArgGlnTyrThrSerLeu 120
 QY 380 CTTCGGGAAAATGAATTCAGAGAGAGAGATGAAGTGTGGCAGGTATCATAGACCCAGA 439
 DB 121 LeuGlyLysMetLysSerGlnGluAspGluValTrpGlnValIleIleGlyAlaArg 140
 QY 440 GCGAGATGACTCTCAAAACCAAGAGTGAAGTGTGGCAGGTATCATAGACCCAGA 499
 DB 141 AlaGluMetThrSerLysHisGlnGluTyrLeuLysLeuGluThrThrTrpMetThrAla 160
 QY 500 GTTGGCTTTTCAGAGATGCGACAGAAAGCTGATATCAAACTGGCGAGATAGAGCCTCT 559

DB 161 ValGlyLeuSerGluMetAlaAlaGluAlaAlaTyrGlnThrGlyAlaAspGlnAlaSer 180
 QY 560 ATAAACCCGACAGAAATCATCATTCAGCTGTGAACTGCAGCGTGAAGAGTGCACACAGCTC 619
 DB 181 IleThrAlaArgAsnHisIleGlnLeuValLysLeuGlnValGluGluValHisGlnLeu 200
 QY 620 TCCCGGAAACAGACAAACCAAGCTGGCGAGAGCACATATACAGAGCTCCGTCAGAAACA 679
 DB 201 SerArgLysAlaGluThrThrLysLeuAlaGluAlaGlnIleGlnGluLeuArgGlnLysThr 220
 QY 680 CAGAGAGAGAGGAGAGGAGCGGCTGAGTGCAGACAGAGAGGCTTACCTGCGTGAAGAT 736
 DB 221 Gln 739
 RESULT 2
 AAU78447
 ID AAU78447 standard; Protein; 239 AA.
 XX
 AC AAU78447;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Inhibitor of apoptosis (IAP) protein Smac.
 XX
 DE Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1; BIR2;
 KW Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;
 KW neoplastic cell; tumour.
 XX
 KW Homo sapiens.
 OS
 XX
 PN WO200216418-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26492.
 XX
 PR 24-AUG-2000; 2000US-227735P.
 XX
 PA (UVE-) UNIV JEFFERSON THOMAS.
 XX
 PI Alnemri ES;
 XX
 DR MPI: 2002-304115/34.
 DR N-PSDB; ABR15451.
 PT Novel Smac peptides and polynucleotides encoding the peptides, useful
 PT for stimulating apoptosis in neoplastic or tumour cell which
 PT overexpresses inhibitor of caspase, and for identifying apoptosis
 PT modulating compounds
 XX
 PS Claim 36; Page 73-74; 78pp; English.
 XX
 CC The invention relates to an isolated Smac peptide or polypeptide (I)
 CC and an isolated nucleic acid (II) encoding (I). Also described is a
 CC method of identifying a compound that inhibits apoptosis, comprising:
 CC (a) separately contacting several cell populations expressing a
 CC cytosolic Smac (a Smac isoform that begins with MKSDPYF sequence,
 CC replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
 CC and residues 56-60 of (I)) and an inhibitor of BID (Bcl2 interacting
 CC domain) with a compound to be tested for apoptotic inhibiting activity;
 CC (b) incubating the cell populations with a direct stimulus of the cell
 CC death pathway; and (c) measuring the specific apoptotic activity of the
 CC cell populations, where inhibition of the specific apoptotic activity is
 CC indicative that the compound is an inhibitor of apoptosis. (I) and (II)
 CC are useful for inducing apoptosis in a cell. The Smac polypeptide and
 CC polynucleotide are useful for stimulating apoptosis in a neoplastic or
 CC tumour cell which overexpresses an inhibitor of caspase, where the
 CC inhibitor inhibits activation or activity of caspase-3, caspase-7 or
 CC caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
 CC (I) is useful for identifying an inhibitor or enhancer of a caspase-
 CC mediated apoptosis which involves contacting a cell transformed or
 CC transfected with a vector expressing (I) with a candidate inhibitor or
 CC candidate enhancer; and detecting cell viability, where an increase in

THIS PAGE BLANK (USPTO)

submitted to the EMBL Data Library, October 1995

A:Reference number: Z19700

A:Accession: T23157

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1133 <M12>

A:Cross-references: EMBL:Z66514; PIDN:CA91344.1; GSPDB:GN00021; CESP:F59A2.6

A:Experimental source: clone K01A11

C:Genetics:

A:Gene: CESP:F59A2.6

A:Map position: 3

A:Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1

Alignment Scores:

Pred. No.:	0.197	Length:	1133
Score:	116.50	Matches:	54
Percent Similarity:	42.73%	Conservative:	43
Best Local Similarity:	23.79%	Mismatches:	75
Query Match:	4.76%	Indels:	55
DB:	2	Gaps:	9

US-09-939-293-1 (1-1358) x T22976 (1-1133)

```

QY 197 GCACAGAAATCAGAGCCCTCATCTCCTTAGTACGATGAGACAGACAGTCTCT 256
    |||||
DB 436 AAGAGLysArgSerSerGluGluGluGluThrAlaAsnGluMetValArgSerLeuThr 455
QY 257 TTGGTAACAGATAGCAGCTTAC- 280
    |||||
DB 456 AAThrLeuGluAsnSerAsnSerGluThrGluLeuLeuLysGluLysLeuThrLeu 475
QY 281 -----TTTCTCTCAGACACATATGCTGATGAGACCTTACT- 322
    |||||
DB 476 AspLysGluLeuGluAlaArgGlnGlnThrGluLysAlaLeuThrGluLysLeuVal 495
QY 323 -----GAATATCTAGAGCTGTTTATCTTACT- 352
    |||||
DB 496 LeuThrThrSerLeuAlaGluLysGluGlnGlnThrAlaGlnGlnGlnGlnGlnThr 515
QY 353 TCTCTTTACCGCAATATACAAATTACTTGGGAAATGATGATGAGACAGACAGTAA 412
    |||||
DB 516 GlnIleTyrGlnMet-----GluValGluLysGluLysValGlu 529
QY 413 GGTGTGGAGGTATCATAGAGCCAGAGCTGAGTACCTTCAAAACACAGAGTACTG 472
    |||||
DB 530 LeuValLysValGlnLeuGlnGlnAlaAlaGlnSerSerSerAlaGluGluAlaLeu 549
QY 473 AAGCTGGAACCACTTGATGACTGCA-----GTTGCTCTTTCAGAGATGCGACAGAA 526
    |||||
DB 550 ArgAlaGluLysGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 569
QY 527 GCTGCATATCAAACTGCGCCAGATCAGCCCTTATACCCGCAAGATAC----- 577
    |||||
DB 570 AlaLeuAsnSerLeuLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 589
QY 578 -----ATTCAGCTGTGGAACCTGAGTGGAGAGAGTGCACCGCTC 619
    |||||
DB 590 ValGluLysGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 609
QY 620 TCCCGGAAAGCAAACTGCGCAGAGACAGACAGTGAAGAGCTCCGTCAGAAACA 679
    |||||
DB 610 SerSerSerValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 628
QY 680 CAGGAGGAAGGAGGAGGAGG-----GCTAGCTGCGACGAG 715
    |||||
DB 629 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 715
QY 716 GAG-----GCCTACCTGGCTGAG 733
    |||||
DB 649 GlnThrAlaGlnIleGlnGlnGln 655

```

RESULT 4
S02041

dystrophin, muscle - chicken
N:Alternate names: duchenne muscular dystrophy protein
C:Species: Gallus gallus (chicken)
C:Date: 07-Sep-1990 #sequence revision 27-Jun-1994 #text change 16-Jul-1999
C:Accession: S02041; S02013; S71487
R:Lemaire, C.; Heilig, R.; Mandel, J.L.
Nucleic Acids Res. 16, 11815-11816, 1988
A:Title: Nucleotide sequence of chicken dystrophin cDNA.
A:Reference number: S02041; MUID:89098331; PMID:3062582
A:Accession: S02041
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3660 <LEM>
A:Cross-references: EMBL:X13369; NID:963369; PIDN:CA91344.1; PID:963370
A:Note: 1869-His, 1885-Arg, and sequences lacking 1171-Met were also found
R:Lemaire, C.; Heilig, R.; Mandel, J.L.
EMBO J. 7, 4157-4162, 1988
A:Title: The chicken dystrophin cDNA: striking conservation of the C-terminal coding
A:Reference number: S02013; MUID:89210800; PMID:3072195
A:Accession: S02013
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3573; 'HA', 3576-3660 <LEM2>
R:Heilig, R.; Lemaire, C.; Mandel, J.L.
Nucleic Acids Res. 15, 9129-9142, 1987
A:Title: A 230kb cosmid walk in the Duchenne muscular dystrophy gene: detection
A:Reference number: S09071; MUID:88067745; PMID:2825128
A:Accession: S71487
A:Molecule type: DNA
A:Residues: 222-281 <HEI>
C:Comment: Dystrophin is proposed to play a role in anchoring the cytoskeleton to li
C:Comment: Defects in dystrophin are responsible for the Duchenne/Becker muscular dy
C:Superfamily: dystrophin; alpha-actinin; actin-binding domain; leucine zipper; membra
C:Keywords: actin binding; calmodulin binding; cytoskeleton; leucine zipper; membra
F:18-237/Domain: alpha-actinin actin-binding domain homology <ACT>
F:253-327/Region: hinge
F:340-449/Domain: spectrin/dystrophin repeat homology <SP01>
F:450-558/Domain: spectrin/dystrophin repeat homology <SP02>
F:560-669/Domain: spectrin/dystrophin repeat homology <SP03>
F:670-719/Region: hinge
F:720-830/Domain: spectrin/dystrophin repeat homology <SP04>
F:838-936/Domain: spectrin/dystrophin repeat homology <SP05>
F:940-1047/Domain: spectrin/dystrophin repeat homology <SP06>
F:1049-1156/Domain: spectrin/dystrophin repeat homology <SP07>
F:1157-1265/Domain: spectrin/dystrophin repeat homology <SP08>
F:1267-1369/Domain: spectrin/dystrophin repeat homology <SP09>
F:1376-1479/Domain: spectrin/dystrophin repeat homology <SP10>
F:1480-1570/Domain: spectrin/dystrophin repeat homology <SP11>
F:1572-1678/Domain: spectrin/dystrophin repeat homology <SP12>
F:1680-1784/Domain: spectrin/dystrophin repeat homology <SP13>
F:1787-1877/Domain: spectrin/dystrophin repeat homology <SP14>
F:1878-1984/Domain: spectrin/dystrophin repeat homology <SP15>
F:1986-2103/Domain: spectrin/dystrophin repeat homology <SP16>
F:2105-2211/Domain: spectrin/dystrophin repeat homology <SP17>
F:2213-2319/Domain: spectrin/dystrophin repeat homology <SP18>
F:2323-2419/Domain: spectrin/dystrophin repeat homology <SP19>
F:2420-2467/Region: hinge
F:2468-2574/Domain: spectrin/dystrophin repeat homology <SP20>
F:2576-2683/Domain: spectrin/dystrophin repeat homology <SP21>
F:2685-2799/Domain: spectrin/dystrophin repeat homology <SP22>
F:2801-2928/Domain: spectrin/dystrophin repeat homology <SP23>
F:2930-3037/Domain: spectrin/dystrophin repeat homology <SP24>
F:3038-3075/Region: hinge
F:3052-3089/Domain: MW repeat homology <MW1>
F:3079-3357/Region: cysteine-rich
F:3481-3502/Region: leucine zipper motif
F:3547-3568/Region: leucine zipper motif

Alignment Scores:

Pred. No.:	0.524	Length:	3660
Score:	112.00 <td>Matches:</td> <td>56</td>	Matches:	56
Percent Similarity:	40.54% <td>Conservative:</td> <td>49</td>	Conservative:	49
Best Local Similarity:	21.62% <td>Mismatches:</td> <td>112</td>	Mismatches:	112

BEST AVAILABLE COPY
THIS PAGE BLANK (USPTO)

```
Query Match: 4.588 Indels: 42
DB: 1 Caps: 8
```

US-09-939-293-1 (1-1358) x S02041 (1-3660)

QY	5	TCGCGGCGTGCACAAATGGGGCGCTCGAAGAGTGGCGTGTCCGACAGCTAACTCATTC	64
Db	1163	SetleuarglyasapleuSerGluMetHisGluTrpIleThrGlnAlaIleuGluIuTr	1185
QY	65	TTACAGCTACAGACAGTGTGTGTGTCTTCGTGTGGCTAACTTTAAAGAGCGGTTC	124
Db	1183	LeuGlu-----ArgAspHeGluTrpIleTrpPro	1192
QY	125	TCGAATTTGATAAGACCACTGGCAAAACTGTGACAGTGGCTTTGAGTAACCTGTGT	184
Db	1193	GluGluLeuGlnIys-----	1197
QY	185	GCGGTTCCTATTGTCAGAGAAATCAGACCTCATTCCTCCTAGTGAAGCATGATGAG	244
Db	1198	AlaValGluGluIuLeuIysArgAlaIysGlnAlaMetGlnIysGluAla-----	1214
QY	245	AGAGCAGTCTTTGGTAAACAGATAGACACCTTACCTTCTCTCCACAGCACATATGCG	304
Db	1215	--LysValIysLeuIleThrAspSerValAsnAsnPhelIleAlaIysAlaProProIa	1233
QY	305	TTGATTGAAGCATATTCTGAATTAATTAAGCTGTTTATTACCTTAACCTTCTTTACCGA	364
Db	1234	AlaAsnGlnAlaLeuIysIysGluLeuAspValLeu-----IleThrSerTyrGlnArg	1251
QY	365	CAATATACAAATTTACTTGGGAAAATGAATTCAGAGGAGGAAGATGAAGTGGCAGGTG	424
Db	1252	LeuIysSerArgLeuAsnGlnIysScyIysTrpIleGlu-----GluValTrp-----	1267
QY	425	ATCATAGGAGCCAGAGACTAGATGACTTCAAAAACCAAGACTTGAAGCTGGAAAC	484
Db	1268	-----AlaCysTrpHisGluLeuLeuSer-----TyrLeuAspAlaGluIuAsn	1281
QY	485	ACTTGGATGACAGTCAATTTGCTTTTCACAGATGGCAGGACGAAGCGCATTTCAAACTGCG	544
Db	1282	LysTrpLeuAsnGluValGluIuLeuIysLeuIysAlaTrpGlu--AsnIleGlnIyGly	1300
QY	545	GCAGATCAGGCTCTAATACCGCCGACGAATCATTCAGTGGTGAACCTCAGGTGGAA	604
Db	1301	AlaGluGlnIleSerGluSerIleuAspSerIleuGlnAlaGluMetArgHisProGluAsp	1320
QY	605	GAGGTGCACCACTCTCCCGGAAGACAGAAACCAAGCTGGCAGAGACACAGATGAAGAG	664
Db	1321	AsnArgAsnGlnIleArgGluLeuAlaGlnIleThrIleuThrAspGlyIleLeuAspGlu	1340
QY	665	CTCCGTCACAAAAACAGAGAGAGAGGAGGCGGGCTGAGTGGAGACAGAGCGCTTAC	724
Db	1341	LeuIleAsnGlnIysLeuGlnIuIysPheAsnThrArgTrpGluGluLeuGlnGlnGluAla	1360
QY	725	CTGCGTGAGAGATTGAGGCGCTGAGACACACCGCCGTCTCCCACTCAGTGGGAA	781
Db	1361	ValArgArgGlnIysSerLeuGlnIuSerIleGlnSerAlaGlnGluTrpAspIys	1379

RESULT 5

N-methyl-D-aspartate receptor chain NMDAR2D-2 - rat
C.Species: *Rattus norvegicus* (Norway rat)
C.Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
C.Accession: I78557, I58158; D45219
R.Monier, H.; Burnashev, N.; Laurie, D.J.; Sakmann, B.; Seeburg, P.H.
Neuron 12, 529-540, 1994
A.Title: Developmental and regional expression in the rat brain and functional properties of NR2B subunit of the NMDA receptor
A.Reference number: I58158; MUID:94206533; PMID:7512349
A.Accession: I78557
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-1323 <RES>
A.Cross-references: GB:LI3612; NID:g469068; PIDN:AAC37647.1; PID:g469069
A.Accession: I58158

A:Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-66, 'V', 68-1323 <RE2>
A:Cross-references: GB:L31611; NID:9469066; PIDN:AA037646.1; PID:9469067
A:Smith, T.: Moriyoshi, K.; Sugihar, H.; Sakurada, K.; Kadotani, H.; Yokoi, M.; Akai,
J. Biol. Chem. 268, 2836-2843, 1993
A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor
A:Reference number: A45219; MUID:93155102; PMID:8428958
A:Accession: D45219
A:Status: preliminary: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1265-1323 <ISH>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP.124265)
A:Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology
F:451-879/Domain: glutamate receptor homology <GRH>

Alignment Scores:

Pred. No.:	0.569	Length:	1322
Score:	111.00	Matches:	56
Percent Similarity:	31.37%	Conservative:	8
Best Local Similarity:	27.45%	Mismatches:	73
Query Match:	4.60%	Indels:	67
DB:	2	Gaps:	9

US-09-939-293-1 (1-1358) x I78557 (1-1323)

QY	915	GCAGACCTGAGACACACAGAGGACATCACAGCTCACAAAGCGCTCGGCTGATTGGCCA	856
Db	1004	AlaLeValaIrgIuIngluIngluProthIngluProProAlaIgluAlaPheProClyPhePro	1022
QY	855	GGCGAGAGACCTGGCGCTCTTCTCGGTCACAGACAGTCATGCCAACCTGGGACAGG	796
Db	1024	SerProProAlaProAlaProAlaAlaAlaAlaAlaAlaValaIgluProProLeucysArgLeu	1043
QY	795	GCA-----TCGGCCCTCGCTTCCCGACAGTAGG-----	766
Db	1044	AlaPheGluAspIuSerProProAlaProSerAlaArgTrpProArgSerAspGluSer	1063
QY	765	-----GGACAGACGGCAGTGTGCTGCAGGCCCTCAATCCACAGCAGTAGGCC	718
Db	1064	GlnProLeuLeuIgluIgluAla-----GlyGlyProSerAlaGlyAlaProThal	1083
QY	717	TCCTGCTCGACATCAGCCGCGCTCTCC-----CTTCCCTCTGTGT	676
Db	1081	AspProProGArgArgAlaAlaProProGArgAlaIgluAspLeuGluProSer	1103
QY	675	TTCTGAC-----GGAGCTCTTATCTGTGCTCTTCTGCCAC	640
Db	1101	oserAspSerGluAspSerGluSerLeuIgluAlaSerLeuGlyIgluGluProTr	1121
QY	639	TTTGTTTCTG-----CTTCCGGGAGAGCTGTGCACCT-----	606
Db	1121	ptrPheAlaAspPheProTyrTrpTyrAlaGluArgLeuGlyProProGlyAlaGly	1141
QY	605	-----CTTCCACCTGC	595
Db	1141	rTrpSerValAspIysLeuGlyIgluTrpArgAlaGlySerTrpAspTyrLeuProProArg	1161
QY	594	AGTTTACCAGCT-----GAATGATTTCTGGCGGTTATAGAGGCTGATCT	547
Db	1161	gGlyIgluProAlaTrpHisCysArgHisCysAlaSerLeuGluIleLeuProProArg	1181
QY	546	GGCGCAGTTGATATAGACACCTCTGCTGCCACATCTGAAGAAGCCAACTGCACTATCCA	487
Db	1181	gHis-----LeuSerCysSerHisAspGlyLeuAspGlyIgluTr	1194
QY	486	GTGGTTTCCA	477
Db	1194	ptrPheAlaPro	1197
RESULT 6			
145219			

BEST AVAILABLE COPY
THIS PAGE IS BLANK